



UNIVERSIDAD  
POLITECNICA  
DE VALENCIA



MÁSTER INTERNACIONAL EN MEJORA GENÉTICA ANIMAL Y  
BIOTECNOLOGÍA DE LA REPRODUCCIÓN

**ANÁLISIS GENÓMICO DE LÍNEAS DIVERGENTES  
SELECCIONADAS PARA CAPACIDAD UTERINA  
EN CONEJO**

Tesis de Máster

Valencia, Septiembre 2016

**Bolívar Samuel Sosa Madrid**

Directora:

María Antonia Santacreu Jerez

## **AGRADECIMIENTOS**

Primero que todo, yo quiero agradecer a Dios por mi vida y su guía en mi caminar a través de este mundo. Igualmente, a mi familia por ayudarme a alcanzar mis metas personales.

Yo quiero agradecer a mis amigos y compañeros su apoyo. Igualmente, por la colaboración en el uso de herramientas bioinformáticas, a Jesús, Noelia, John, Mara y Roger. Además, quiero dar las gracias a José Cedano e Ion por su ayuda sobre ideas científicas y planteamientos de problemas de investigación. También, a mi tutora María Antonia Santacreu por su apoyo y consejo en el desarrollo de mi tesis final de máster. Agradezco a Agustín por inculcarme un pensamiento crítico. A Ahmed, Manuel, Pilar, Verónica, Marina y Fede por su apoyo en el trabajo diario.

Finalmente, a todos aquellos que de alguna forma colaboraron con el desarrollo de este documento presentado para su evaluación.

## CONTENIDO

ABSTRACT .....	viii
RESUMEN.....	x
RESUM.....	xii
INTRODUCCIÓN .....	1
1. Importancia económica del tamaño de camada en conejo.....	1
2. Tamaño de camada y sus componentes.....	2
2.1. Tamaño de camada.....	2
2.2. Tasa de ovulación .....	4
2.3. Supervivencia prenatal .....	5
2.4. Capacidad uterina .....	6
3. Experimentos de selección para mejorar el tamaño de camada.....	7
4. Genómica de los caracteres reproductivos .....	10
4.1. Frecuencia Alélica y Deriva Génica .....	11
4.2. Genes candidatos para capacidad uterina en conejo .....	15
4.3. Estudio de asociación del genoma completo (GWAS).....	17
OBJETIVOS .....	23
ARTICULO 1: ESTRUCTURA GENÉTICA DE DOS LÍNEAS DIVERGENTES SELECCIONADAS PARA CAPACIDAD UTERINA EN CONEJO.....	24
Resumen .....	24
Introducción .....	25
Material y Métodos.....	26
Resultados.....	34
Discusión.....	45
Conclusiones.....	49
Agradecimientos .....	49
Referencias Bibliográficas.....	49
ARTÍCULO 2: ESTUDIO DE ASOCIACIÓN DEL GENOMA COMPLETO Y HUELLAS DE SELECCIÓN EN LINEAS DIVERGENTES SELECCIONADAS PARA CAPACIDAD UTERINA EN CONEJO .....	53
Resumen .....	53
Introducción .....	54

Material y Métodos.....	55
Resultados.....	62
Discusión.....	72
Conclusiones e Implicaciones.....	75
Agradecimientos.....	76
Referencias Bibliográficas.....	76
DISCUSIÓN GENERAL Y CONCLUSIONES.....	85
REFERENCIAS BIBLIOGRÁFICAS.....	88

# ÍNDICES DE TABLAS

## Introducción

<b>Tabla 1.</b> Tamaño de camada al nacimiento en diversas razas/líneas con diferentes partos.....	3
<b>Tabla 2.</b> Respuesta a la selección en tamaño de camada en diversas especies prolíficas.....	8

## Artículo 1

<b>Tabla 1.</b> Número de QTNs por grupo de estructura genética.....	27
<b>Tabla 2.</b> Clasificación de SNP de acuerdo con la asociación .....	31
<b>Tabla 3.</b> Datos descriptivos de QTN-Fijados.....	36
<b>Tabla 4.</b> Resumen de los resultados promedios por grupo utilizando la estrategia 2. ....	38
<b>Tabla 5.</b> Parámetros descriptivos del error tipo I (T1E) usando la estrategia 2 (líneas divergentes) .....	38
<b>Tabla 6.</b> Parámetros descriptivos de la potencia de detección de asociación de SNP usando la estrategia 2 (líneas divergentes) .....	39
<b>Tabla 7.</b> Parámetros descriptivos del valor predictivo positivo usando la estrategia 2 (líneas divergentes) .....	39
<b>Tabla 8.</b> Parámetros de estimación del efecto de la deriva génica en los tres escenarios para las líneas divergentes.....	42

## Artículo 2

<b>Tabla 1.</b> Análisis descriptivos del tamaño de camada y sus componentes. ....	57
<b>Tabla 2.</b> Número de SNPs excluidos por diferentes criterios .....	58
<b>Tabla 3.</b> Las varianzas y el valor pi ( $\pi$ ) para el método de Bayes B.....	60
<b>Tabla 4.</b> Regiones de QTL asociadas con el tamaño de camada y número de nacidos vivos en conejo por Bayes B.....	64
<b>Tabla 5.</b> Regiones de QTL asociadas con la tasa de ovulación en conejo por Bayes B. ....	65

<b>Tabla 6.</b> Regiones de QTL asociadas con el número de embriones implantados en conejo por Bayes B. ....	66
<b>Tabla 7.</b> Número y características de los “clúster” obtenidos a partir de los índices de diferenciación genética ( $F_{st}$ ) entre las poblaciones estudiadas, la línea alta de capacidad uterina (Alta) - la línea de referencia (línea V), y la línea baja de capacidad uterina (Baja) y la línea de referencia (línea V) .....	69
<b>Tabla 8.</b> Número de genes candidatos, principales genes candidatos y funciones biológicas de los “clúster” obtenidos a partir de los índices de diferenciación genética ( $F_{st}$ ) entre la línea baja de capacidad uterina (Baja) y la línea de referencia (línea V). ....	71

# ÍNDICES DE FIGURAS

## Introducción

**Figura 1.** Árbol de una Coalescencia de una muestra de  $n= 5$  unidades. .... 13

## Artículo 1

**Figura 1.** Esquema de simulación de la selección divergente de capacidad uterina...29

**Figura 2.** Ejemplo de creación de una matriz de confusión y un gráfico ROC.. .....33

**Figura 3.** Relación entre la respuesta a la selección versus la varianza aditiva inicial..  
.....35

**Figura 4.** Número de QTNs fijados por cada grupo y estrategia.. .....37

**Figura 5.** Violin-Plot. (a) error tipo I, (b) potencia de detección, (c) valor predictivo positivo..... 40

**Figura 6.** Gráfico ROC utilizando resultados predictivos de la matriz de confusión para calcular la potencia y el error tipo I. .... 41

**Figura 7.** Gráficos de las medias del verdadero valor mejorante (TBV) por generación en diferentes escenarios de deriva génica y selección..... 44

## Artículo 2

**Figura 1.** Gráfico de Escalado Multidimensional (MDS). .... 63

**Figura 2.** Desequilibrio de ligamiento en ventanas relevantes del Cromosoma17 ....67

**Figura 3.** Manhattan-Plot de los valores  $F_{st}$  obtenidos de la línea baja de UC con la línea de referencia (línea V) ..... 70

## ABSTRACT

Litter size is an economically important trait in rabbits. Maternal lines are usually selected for litter size and they are used in a three-way crossover to produce rabbit meat. Low response obtained after selection for litter size in closed populations led to the search for alternative methods of selection for improving more efficiently litter size. Selection for main components of litter size (ovulation rate and prenatal survival) has been proposed as a way to improve litter size.

An experiment of divergent selection for uterine capacity was carried out in the Animal Breeding Unit of Polytechnic University of Valencia. Uterine capacity was defined as the maximum number of foetuses that a female is able to support up to birth when ovulation rate is not a limiting factor. Two divergent lines, one to increase uterine capacity (High Line) and one to decrease (Low Line) were selected during ten generations. Uterine capacity was assessed as litter size in unilaterally ovariectomized (ULO) does. After 10 generations, selection was relaxed for six generations. At the beginning of divergent selection process, embryos were vitrified from base population (Line V). In the generation 10, embryos were thawed and a contemporary reference population was established. In the 11<sup>th</sup> and 12<sup>th</sup> generations, blood samples were frozen from 90, 70 and 30 intact females from High, Low and reference line, respectively. Nowadays, the first SNPchip for genotyping in rabbits, 200K Affymetrix Axiom OrcunSNP Array, is available. Thus, main objective was to detect genomic areas associated with litter size in rabbits by GWAS and selection footprints in a divergent selection experiment for uterine capacity (Article 2). In addition, another objective was to study different scenarios of genetic structure for uterine capacity in order to detect positive associations by simulation. The drift genetic effect was also studied (Article 1).

Results using 160 animals of divergent lines had advantages over using the same number of animals from only one line. A genetic structure of 10,010 QTN would have a greater positive predictive value. Genetic drift represented up to 14.52% of the selection response, but it was responsible that up to 50% QTNs fix. GWAS analysis showed four genomic regions associated with litter size (LS). One, five and four associated regions were found for ovulation rate (OR), implanted embryos (IE) and number born alive (NBA), respectively. These genomics regions were considered as putative QTLs. Putative QTLs were not detected for embryo, fetal and prenatal survival (ES, ES and SP, respectively) and number born dead (NBD). The regions of chromosome 17 explained 20.62% of the genomic variance of LS, being themselves the most

relevant genomic region. The ERO1A gene, reported on expression analyses in these rabbit divergent lines, was found within the QTL on chromosome 17. In addition, new genomic areas were found. Their biological functions are related to the activity of prostanoid receptors, cytoplasmic components and to a negative regulation of cell proliferation. Main genes were BMP4, PTGDR ANKH and CDKN3. On the other hand, Fst were not found between the divergent lines, but six clusters were found between the low line of uterine capacity and the line V. Thus, our findings provide a new starting point for further genomic studies, in order to unravel the genetic basis of litter size and its components in rabbits. These could also provide useful information for future research on other selection strategies or the application of genomic selection in rabbit commercial populations.

**Keywords:** Error Type I, statistical power, simulation, uterine capacity, GWAS, litter size, Fst, selection footprints, divergence.

## RESUMEN

El tamaño de camada es un carácter económicamente importante en conejo y es el objetivo de selección en las líneas maternas que se utilizan en un cruce a tres vías para producir el gazapo de carne. Sin embargo, la respuesta a la selección por tamaño de camada ha sido baja. Se han propuesto alternativas para mejorar el tamaño de camada de forma más eficaz y una de ellas es la mejora de los componentes del tamaño de camada.

En la Unidad de Mejora Genética de la Universidad Politécnica de Valencia se llevó a cabo un experimento de selección divergente por capacidad uterina. La capacidad uterina, que es la supervivencia prenatal que depende de la hembra cuando el cuerno uterino está atestado de óvulos, es uno de los principales componentes del tamaño de camada. Durante diez generaciones se seleccionaron dos líneas, una para aumentar la capacidad uterina (Línea Alta) y otra para disminuirla (Línea Baja). La capacidad uterina se estimó como el número de nacidos totales (LS) en hembra unilateralmente ovariectomizadas. Después de 10 generaciones, se relajó la selección durante seis generaciones. Al principio del proceso de selección divergente, se vitrificaron embriones de la línea de la que procedían los animales de la población base, la línea V. En la generación 10 de capacidad uterina, se desvitrificaron los embriones y se constituyó una población de referencia coetánea con las generaciones 11 y 12 del experimento de selección de capacidad uterina. En las generaciones 11 y 12 se congelaron muestras de sangre de 90 hembras intactas de la línea Alta, 70 de la Línea Baja y 30 de la línea de referencia. Actualmente se dispone del primer SNP-chip para genotipado en conejo, 200K Affymetrix Axiom OrcunSNP Array. Por consiguiente, el objetivo principal del trabajo fue detectar áreas genómicas asociadas al tamaño de camada en conejo mediante un GWAS y las huellas de la selección a partir de las muestras de sangre del experimento de selección divergente por capacidad uterina (artículo 2). Además, otro objetivo de la investigación fue estudiar cómo afectaban diferentes escenarios de estructura genética del carácter a la detección de asociaciones positivas mediante la simulación de este experimento de selección divergente. También se estudió el efecto de la deriva mediante simulación (artículo 1).

Los hallazgos del estudio de simulación muestran que utilizar 160 animales de las líneas divergentes presenta ventajas sobre utilizar sólo una línea y que una estructura genética de 10,010 QTNs tendría un mayor valor predictivo positivo respecto a estructuras genéticas con menor número de QTNs (1,010 y 110). La deriva génica representa hasta un 14.52% de la respuesta de la selección, pero es responsable hasta del 50% de los QTNs fijados. En el GWAS

se encontraron 4 regiones genómicas asociadas al carácter tamaño de camada (LS). Para la tasa de ovulación (OR), el número de embriones implantados (IE) y el número de nacidos vivos (NBA) se encontraron 1, 5 y 4 regiones asociadas, respectivamente. Estas regiones genómicas fueron consideradas como QTLs putativos. En el caso de las supervivencias, embrionaria, fetal y prenatal (ES, FS, PS, respectivamente) y el número de nacidos muertos (NBD) no se detectó ningún QTL putativo. Las regiones del cromosoma 17 representan el 20.62 % de la varianza genética de LS, siendo la región genómica más relevante. El gen ERO1A, reportado en un análisis de expresión en estas líneas divergentes de conejo, se encontró dentro de un QTL del cromosoma 17. Además, se encontraron nuevas áreas genómicas asociadas, cuyas funciones biológicas tienen que ver con la actividad de receptores de prostanoides, los componentes asociados al citoplasma y la regulación negativa de la proliferación celular. Entre estas tenemos: BMP4, PTGDR, ANKH y CDKN3. Por otro lado, no se encontró ningún “clúster” de  $F_{st}$  entre las líneas divergentes, sin embargo se encontraron seis “clúster” entre la línea baja y la línea V. Estos hallazgos ofrecen un nuevo punto de partida para estudios genómicos posteriores, a fin de desentrañar la base genética del tamaño de camada y sus componentes en el conejo. Además, podrían proveer información útil para futuras investigaciones sobre otras estrategias de selección o la aplicación de selección genómica en poblaciones comerciales de conejo.

**Palabras Claves:** Error Tipo I, análisis de potencia, simulación, capacidad uterina, GWAS, tamaño de camada,  $F_{st}$ , huellas de selección, divergencia.

## RESUM

La grandària de ventrada és un caràcter econòmicament important en conill i és l'objectiu de selecció en les línies maternes que s'utilitzen en un encreuament a tres vies per produir el llogó de carn. No obstant això, la resposta a la selecció per grandària de ventrada ha estat baixa. S'han proposat alternatives per millorar la grandària de ventrada de forma més eficaç i una d'elles és la millora dels components de la grandària de ventrada.

En la Unitat de Millora Genètica de la Universitat Politècnica de València es va dur a terme un experiment de selecció divergent per capacitat uterina. La capacitat uterina, que és la supervivència prenatal que depèn de la femella quan la banya uterina aquesta atestat d'òvuls, és un dels principals components de la grandària de ventrada. Durant deu generacions es van seleccionar dues línies, una per augmentar la capacitat uterina (Línia Alta) i una altra per disminuir-la (Línia Baixa). La capacitat uterina es va estimar com el nombre de nascuts totals (LS) en femelles unilateralment ovariectomizadas. Després de 10 generacions, es va relaxar la selecció durant sis generacions. Al principi del procés de selecció divergent, es vitrificaron embrions de la línia de la qual procedien els animals de la població basi, la línia V. En la generació 10 de capacitat uterina, es desvitrificaron els embrions i es va constituir una població de referència coetània amb les generacions 11 i 12 de l'experiment de selecció de capacitat uterina. En les generacions 11 i 12 es van congelar mostres de sang de 90 femelles intactes de la línia Alta, 70 de la Línia Baixa i 30 de la línia de referència. Actualment es disposa del primer SNP-xip per genotipado en conill, 200K Affymetrix Axiom OrcunSNP Array. Per tant, l'objectiu principal del treball va ser detectar àrees genòmiques associades a la grandària de ventrada en conill mitjançant un GWAS i les petjades de la selecció a partir de les mostres de sang de l'experiment de selecció divergent per capacitat uterina (article 2). A més, un altre objectiu de la recerca va ser estudiar com afectaven diferents escenaris d'estructura genètica del caràcter a la detecció d'associacions positives mitjançant la simulació d'aquest experiment de selecció divergent. També es va estudiar l'efecte de la deriva mitjançant simulació (article 1).

Les troballes de l'estudi de simulació mostren que utilitzar 160 animals de les línies divergents presenta avantatges sobre utilitzar només una línia i que una estructura genètica de 10,010 QTNs tindria un major valor predictiu positiu. La deriva gènica representa fins a un 14.52% de la resposta de la selecció, però és responsable fins al 50% de QTN fixats. En el GWAS es van trobar 4 regions genòmiques associades al caràcter grandària de ventrada (LS). Per a la

taxa d'ovulació (OR), el nombre d'embrions implantats (IE) i el nombre de nascuts vius (NBA) es van trobar 1, 5 i 4 regions associades, respectivament. Aquestes regions genòmiques van ser considerades com QTLs putatius. En el cas de les supervivències, embrionària, fetal i prenatal (ÉS, FS, PS, respectivament) i el nombre de nascuts morts (NBD) no es va detectar cap QTL putatiu. Les regions del cromosoma 17 representen el 20.62 % de la variància genètica de LS, sent la regió genòmica més rellevant. El gen ERO1A reportat en anàlisi d'expressió en aquestes línies divergents de conill es va trobar dins d'un QTL del cromosoma 17. A més, es van trobar noves àrees genòmiques associades, les funcions biològiques de les quals té a veure amb l'activitat de receptors de prostanoides, components associats al citoplasma i la regulació negativa de la proliferació cel·lular. Entre aquestes tenim: BMP4, PTGDR, ANKH i CDKN3. D'altra banda, no es van trobar "clúster" de Fst entre les línies divergents, però sis "clúster" van ser trobats entre la línia baixa i la línia. Per tant, les nostres troballes ofereixen un nou punt de partida per a estudis genòmics posteriors, a fi de desentranyar la base genètica de la grandària de ventrada i els seus components en el conill. A més, podrien proveir d'informació útil per a futures recerques sobre altres estratègia de selecció o l'aplicació de selecció genòmica en poblacions comercials de conill.

**Paraules Claus:** Error Tipus I, anàlisi de potència, simulació, capacitat uterina, GWAS, grandària de ventrada, Fst, petjades de selecció, divergència.

# INTRODUCCIÓN

La incorporación de la genómica a la investigación ha producido nuevas fuentes de conocimiento sobre enfermedades hereditarias, rutas metabólicas y sistemas biológicos. En los programas de mejora genética en especies de interés productivo, el uso de la genómica ha ayudado a una mejor predicción del valor genético de un animal en el ganado vacuno (Pryce y Daetwyler, 2012). Sin embargo, la utilización de las herramientas genómicas en especies prolíficas (cerdo, pollo y conejo) ha sido limitada, porque la principal ventaja obtenida por la genómica en las especies no prolíficas (reducir el alto intervalo generacional) no puede ser utilizada con el mismo impacto económico (Blasco y Toro, 2014).

El conejo doméstico (*Oryctolagus cuniculus*) ha sido un animal utilizado tanto para la producción como para la experimentación como modelo animal en humanos y otras especies. Para aumentar la producción de carne, las poblaciones de conejos han sido sometidas a selección artificial para caracteres reproductivos y de crecimiento. En la actualidad existen herramientas genómicas disponibles para mejorar los programas de selección en conejos y para ayudar a comprender la estructura genética de los caracteres de interés económico. Además, el estudio con estas herramientas genómicas en conejo podría servir de modelo para otras especies prolíficas de interés productivo.

## 1. Importancia económica del tamaño de camada en conejo.

Los principales ingresos de las granjas comerciales de conejo son las ventas de animales para matadero. Así pues, el objetivo principal de la granja comercial será producir el mayor número de conejos comerciales con el menor coste posible.

El tamaño de camada es el carácter reproductivo con mayor peso económico en conejos (Cartuche et al., 2014). Al aumentar el tamaño de camada, los costes fijos, que son más importantes que en otras especies, se dividen por más individuos y por lo tanto los costes de conejo por kilo vendido disminuyen. El criterio de selección más utilizado para mejorar el tamaño de camada es el número de gazapos destetados que recoge tanto la capacidad de la hembra para parir un número elevado de gazapos como la capacidad de cuidarlos hasta el momento del destete (Baselga y Blasco, 1989).

Habitualmente, el esquema de selección en conejos se basa en un cruce a tres vías donde las dos líneas maternas habitualmente son seleccionadas por tamaño de camada al destete y la línea paterna es seleccionada por velocidad de crecimiento entre el destete y el sacrificio. (Baselga y Blasco, 1989).

## **2. Tamaño de camada y sus componentes**

### **2.1. Tamaño de camada**

El tamaño de camada al nacimiento es un carácter complejo desde el punto de vista biológico ya que es el resultado de una secuencia de sucesos: ovulación, fecundación y viabilidad de los embriones hasta el momento del parto (Santacreu, 2006). Las medidas más comunes del tamaño de camada son en el parto, al destete y al sacrificio que es cuando alcanzan el peso comercial (63 días para el conejo en España).

En la Tabla 1 se muestran la media y la desviación típica del tamaño de camada al nacimiento de algunos trabajos publicados, observándose diferencias notables entre las medias y una oscilación entre 1.24 y 3.35 de desviación estándar. Estas diferencias entre poblaciones se pueden deber principalmente a diferencias genéticas (razas o líneas que se están estudiando) o a diferencias ambientales como la estación del año, la edad, el orden de parto, el fotoperiodo durante la gestación, el estado fisiológico reproductivo en el momento de la cubrición y el intervalo parto-cubrición. (Fayeye y Ayorinde, 2010; Poujardieu y Théau-Clément, 1995)

En cuanto a los parámetros genéticos, la heredabilidad del tamaño de camada es baja entre 0.03 a 0.19 (Blasco et al., 1995) de acuerdo con lo que ocurre en conejo, porcino y ratón (Clutter et al., 1990; Egena et al., 2012; Ferraz y Eler, 1994; Johnson et al., 1999). Conforme nos alejamos más del momento del parto la heredabilidad del tamaño de camada es más pequeña, ya que la mortalidad de los gazapos está influida en mayor medida por los factores ambientales y en menor medida por los efectos genéticos de la hembra. Las estimas de heredabilidad para el tamaño de camada al destete se encuentran entre 0.01 a 0.12 y se aproximan a cero en el momento del sacrificio (Baselga y Blasco, 1989; García y Baselga, 2002a).

La respuesta a la selección por tamaño de camada ha sido baja, alrededor de 0.01 gazapos por generación en diferentes experimentos (Khalil y Al-Saef, 2008). Por este motivo, se han

propuesto diferentes alternativas para mejorar estos resultados. Una de las alternativas propuesta para mejorar la respuesta en el tamaño de camada ha sido la selección indirecta del tamaño de camada seleccionando por sus componentes. Los principales componentes del tamaño de camada son la tasa de ovulación y la supervivencia prenatal. Se ha estudiado el papel de la madre y el embrión sobre la supervivencia prenatal y se ha visto que la supervivencia prenatal está principalmente determinada por la madre (Mocé et al., 2004).

Tabla 1. Tamaño de camada al nacimiento en diversas razas/líneas con varios partos

Raza /Línea	Media	Desviación Estándar	Coefficiente de Variación (%)	Referencias
Conejo Gabali	5.74	1.24	21.51	(Iraqi et al., 2007)
Líneas Conjuntas (A, V y LP)	11.03	3.35	---	(Ragab et al., 2014)
Línea Botucatu	6.66	3.14	47.10	(Moura et al. 2001)
Línea A*	8.60	2.90	---	(Piles et al., 2006)
Línea Prat*	10.70	2.70	---	(Piles et al., 2006)
Línea V*	10.40	3.10	---	(Piles et al., 2006)
Línea Pannon Ka	9.19	3.11	---	(Nagy et al., 2014)
Neozelandesa Blanca	8.33	2.50	---	(Elmaghraby y Elkholya, 2010)
Neozelandesa Blanca	7.10	2.30	29.00	(Khalil, 1993)
Línea Bauscat	6.48	2.13	32.86	(Khalil et al.1987)
Línea Giza	6.36	2.01	30.96	(Khalil et al. 1987)
Californiana	6.70	2.30	33.00	(Khalil, 1993)
Conejo Danés Blanco	9.53	2.18	---	(Sorensen et al., 2001)

\*Datos provenientes de experimentos con más de dos partos

## 2.2. Tasa de ovulación

Es uno de los principales caracteres que determinan el carácter tamaño de camada. La tasa de ovulación es el número de ovocitos liberados por los ovarios en la ovulación y representa el número máximo de crías que puede tener una hembra en cada camada (Argente, 1996). La tasa de ovulación puede ser estimada post-mortem de diversas maneras: contando el número de ovocitos recuperados mediante la perfusión de los oviductos, o contando el número de cuerpos lúteos presentes en ambos ovarios en un momento dado de la gestación (Lambert et al., 1991; Torrès et al., 1987). No obstante, la técnica de laparoscopia permite una buena estima in vivo de la tasa de ovulación sin repercusiones adversas sobre el tamaño de camada. En efecto, en un estudio se encontró una correlación del 0.99 entre el número de cuerpos lúteos observados por laparoscopia y los puntos de ovulación contados post-mortem por enucleación (Santacreu et al., 1990).

La ovulación de la coneja es inducida por el estímulo de la monta, lo que es una ventaja de manejo en el sistema de producción. Sin embargo, el rechazo o aceptación a la monta por parte de la hembra, parece estar relacionada al número de folículos preovulatorios (García-Ximénez y Vicente, 1992).

La tasa de ovulación también se ve afectada por factores ambientales a pesar de ser inducida por la monta. Entre estos factores están la estación del año (primavera-verano, otoño-invierno), el número de partos, la raza o línea, el peso de la hembra (alimentación) y estado fisiológico (Castellini et al., 2010; Garcia et al., 1984; Garcia y Perez, 1989).

La tasa de ovulación tiene estimas de heredabilidad entre 0.16 a 0.21 en la coneja (Blasco et al., 2005; Bolet y Theau-Clément, 1994; Laborda et al., 2012a), de 0.05 a 0.48 en porcino (Bidanel et al., 1996a; Johnson et al., 1999; Lamberson et al., 1991; Ruíz-Flores y Johnson, 2001) y en ratones de 0.10 a 0.33 (Bradford 1969; Clutter et al. 1990). Aunque la estima depende del estudio, la conclusión principal es que la tasa de ovulación tiene una heredabilidad más alta que el tamaño de camada y que la correlación con el tamaño de camada es baja, por lo tanto no se esperaría una repuesta correlacionada alta en un programa de selección (Laborda et al., 2012b).

### 2.3. Supervivencia prenatal

La tasa de fecundación generalmente es elevada en conejo (Bolet y Theau-Clément, 1994; Santacreu et al., 1996). Sin restricción de la fecundación, la supervivencia durante la gestación juega un papel fundamental en la magnitud del tamaño de la camada al nacimiento. La supervivencia prenatal se estima como el cociente entre el tamaño de camada y la tasa de ovulación (Santacreu, et al. 1990). En término general, la supervivencia prenatal es aproximadamente de un 70% de los óvulos liberados por el ovario tanto en el conejo como en otras especies prolíficas (Santacreu, 2006). Por especie, los valores oscilan entre el 87% y 70% en conejo, entre 80% y 60% en porcino y entre el 95% y 60% en ratón (Blasco et al., 1993; Hernández, 2011; Wilmut et al., 1986).

En conejo, la supervivencia prenatal a lo largo de toda la gestación puede dividirse en dos periodos. A la comprendida entre la fecundación y la implantación se le denomina supervivencia embrionaria y a la comprendida entre la implantación y el parto es conocida como supervivencia fetal (Santacreu, 2006). Las pérdidas estimadas en el periodo embrionario son entre el 10 al 14% y la mortalidad fetal de un 20 a 22% (García y Baselga, 2002b; Santacreu, 1992; Santacreu et al., 2005). Se ha propuesto que los genes que controlan la supervivencia embrionaria y fetal podrían ser diferentes, al igual que en otros componentes del tamaño de camada (Onteru et al., 2011a; Schneider et al., 2012; Spötter y Distl, 2006).

Además de los factores genéticos, la supervivencia prenatal en todas las especies prolíficas productivas está determinada por factores ambientales. Los factores ambientales más importantes son: la estación en la que transcurre la gestación, la edad (asociada también al número de partos) y el estado fisiológico de la hembra (Wilmut et al., 1986). Las investigaciones son contradictorias, lo cual indica que dependiendo de las condiciones del estudio los resultados pueden ser diversos (Baena et al., 2006; Fortun et al., 1994; Foxcroft y Hasnain, 1973; Smitd et al., 1968).

En conejo, las estimaciones de heredabilidad de la supervivencia prenatal están entre 0.08 a 0.23, dependiendo del modelo utilizado para la estimación y la precisión (Bolet et al., 1994; Santacreu, 1992). En cerdo, la estimación de supervivencia prenatal están entre 0.10 a 0.14 (Hernández, 2011; Rosendo et al., 2007) y en ratón se tienen datos de supervivencia a los 17 días de gestación de 0.15 (Gion et al., 1990). En síntesis, una elevada supervivencia prenatal

conduce a un mayor tamaño de camada al nacimiento y consecuentemente a mayores beneficios económicos debido a la relativa alta correlación entre estos dos caracteres (Santacreu, 2006).

#### **2.4. Capacidad uterina**

La capacidad uterina puede definirse de varias formas. La primera definición fue propuesta en porcino por Christenson et al. (1987), como el máximo número de fetos que una hembra puede llevar al término de la gestación cuando no hay una limitación por el número de óvulos producidos. La capacidad uterina se ha relacionado en el periodo fetal con los aspectos físicos medibles en el útero, como su longitud y su volumen o con los factores intrauterinos que permiten una adecuada superficie de intercambio de los nutrientes entre la circulación materna y la fetal (Argente et al., 2003a; Bazer et al., 1969). Una visión más amplia del concepto de capacidad uterina sería la que tendría en cuenta también todos los factores bioquímicos, proteínas y hormonas del ambiente materno del oviducto o del útero que intervienen a lo largo de la gestación en la supervivencia embrionaria y fetal (Freking et al., 2007; Vallet et al., 2002).

Las técnicas que podrían emplearse para conseguir las condiciones de atestamiento uterino para que se pueda expresar la capacidad uterina son: la superovulación, la transferencia de embriones o la ovariectomización unilateral. Las dos primeras técnicas presentan inconvenientes relacionados con la disminución de la supervivencia embrionaria y el tamaño de camada (Badawy et al., 2013; Maertens, 1998; Viudes de Castro et al., 2009).

La ovariectomización unilateral consiste en la extirpación de uno de los ovarios. Al extirpar un ovario se produce un fenómeno de compensación ovárica en el único ovario presente (Blasco et al., 1994; Christenson et al., 1987; Nielsen et al., 1995). Esto significa que el ovario que queda en las conejas ovariectomizadas unilateralmente (hembras ULO) es capaz de ovular la misma cantidad de óvulos que la media de los dos ovarios de una hembra intacta (Blasco et al., 1994). Así, el cuerno uterino adyacente a ese ovario recibe por término medio el doble de embriones que el cuerno uterino de una hembra intacta, debido a que la coneja presenta un útero dúplex y no hay transmigración de embriones entre los dos cuernos uterinos (McNitt et al., 2013). Las conejas ovariectomizadas unilateralmente gestan por término medio el doble de embriones en un mismo cuerno uterino y tienen un elevado tamaño de camada, llegando a un 77 % y 89% respecto al tamaño de camada en las hembras intactas (Argente, 1996; Blasco et al., 1994, 2005). Para estimar la capacidad uterina en porcino, además de quitar un ovario hay que eliminar y ligar el cuerno ipsolateral para evitar la transmigración uterina. La técnica

ha sido ampliamente utilizada en especies prolíficas porque solo debe realizarse una vez y se obtienen datos de todos los partos que tenga la hembra (Clutter et al., 1990; en ratón; Bolet y Theau-Clément, 1994; Santacreu et al., 1996; en conejo; Christenson et al., 1987; en porcino).

La capacidad uterina tiene estimas de heredabilidad de 0.11 a 0.16 en el conejo (Argente et al., 1997; Blasco et al., 2005), de 0.06 a 0.10 en ratones (Ribeiro et al., 1997) y de 0.15 a 0.20 en cerdo, estas últimas estimadas por modelos de simulación (Bennett y Leymaster, 1990). Por lo tanto, una mejora de la capacidad uterina puede suponer una mejora en el número de embriones implantados y en la supervivencia fetal debido a la correlación entre estos caracteres (Argente et al., 1997; Blasco et al., 2005).

Hasta el momento, podemos decir que el tamaño de camada es un carácter complejo determinado por otros caracteres. Los experimentos de selección que se han llevado a cabo por los caracteres componentes del tamaño de camada son escasos. Así, la determinación genética de los caracteres componentes del tamaño de camada es poco conocida, como también las relaciones entre ellos y con el tamaño de camada.

### **3. Experimentos de selección para mejorar el tamaño de camada**

En conejo, como se ha comentado anteriormente, el criterio de selección más utilizado ha sido el número de gazapos al destete, alrededor de 30 días. La respuesta obtenida en líneas comerciales de conejo después de muchas generaciones de selección ha sido alrededor de 0.1 gazapos por generación, la cual es menor a la esperada (Baselga, 2001; Khalil y Al-Saef, 2008). Este resultado es similar al encontrado en los experimentos de selección por el número de nacidos vivos al parto en porcino (Blasco et al., 1993). La selección por los principales componentes del tamaño de camada surge como una alternativa para intentar buscar una forma más eficaz de mejorar el tamaño de camada en las especies prolíficas de interés productivo. Algunos resultados del progreso genético en tamaño de camada y sus componentes en diferentes especies con distintas estrategias de selección (directa o por sus componentes) pueden observarse en la Tabla 2.

Tabla 2. Respuesta a la selección en tamaño de camada en diversas especies prolíficas

Carácter Criterio de Selección	Sentido	Respuesta a la Selección <sup>1</sup>	Nº de Generaciones	Referencia
<b>CONEJO</b>				
Capacidad Uterina	A Favor- En Contra	0.10	12	(Santacreu et al., 2005)
Tamaño de Camada al Destete	A Favor	$0.11 \pm 0.03$	21	(García y Baselga, 2002a)
Tasa de Ovulación	A Favor	$\approx -0.02$	10	(Laborda et al., 2011)
Varianza Residual de Tamaño de Camada	A Favor- En Contra	0.13	5	(Martínez-Álvaro, 2013)
<b>RATÓN</b>				
Índice (Tasa de Ovulación + Supervivencia Prenatal)	A Favor	$0.10 \pm 0.01^a$	21	(Kirby y Nielsen, 1993)
Capacidad Uterina	A Favor	$0.08 \pm 0.01^a$	21	(Kirby y Nielsen, 1993)
Tamaño de Camada al Nacimiento	A Favor	$0.10 \pm 0.01^a$	21	(Kirby y Nielsen, 1993)
<b>PORCINO</b>				
Tasa de Ovulación	A Favor	$0.09 \pm 0.06$	9	(Lamberson et al., 1991)
Índice(Eficiencia Placentaria)	A Favor- En Contra	$0.07 \pm 0.02$	4	(Mesa et al., 2005)
Índice (Ovulación + Supervivencia Embrionaria)	A Favor	$0.22 \pm 0.03^a$	11	(Johnson et al., 1999)
Capacidad Uterina	A Favor	$0.15^a$	11	(Freking et al., 2016)
Supervivencia Prenatal	A Favor	$0.07^b$	4	(Bidanel et al., 1996b)

<sup>1</sup> Respuesta a la selección correlacionada de tamaño de camada. La respuesta fue ajustada y calculada por generación por un solo sentido.

<sup>a</sup> Respuesta estimada como la diferencia entre la población seleccionada y la población control en una generación determinada.

<sup>b</sup> Respuesta correlacionada obtenida a través de un modelo con BLUP y una línea control.

Las respuestas correlacionadas en tamaño de camada cuando se ha seleccionado por tasa de ovulación han sido muy pequeñas e irrelevantes, tanto en conejo como en cerdo y ratón (Bradford, 1969; Laborda et al., 2012a; Lamberson et al., 1991; respectivamente).

Los resultados de selección por supervivencia prenatal en cerdos demuestran que no hay una respuesta relevante en la supervivencia y en el tamaño de camada (Bidanel et al., 1996b). En conejo no se ha realizado ningún experimento de selección por supervivencia prenatal, pero en ratón se han encontrado resultados positivos después de once generaciones de selección. En este experimento se obtuvo un aumento de 0.80 a 0.90 de supervivencia prenatal a los 17 días (Bradford, 1979) pero la respuesta correlacionada en tamaño de camada no fue relevante.

Cuando se ha seleccionado por índices de selección para tasa de ovulación y supervivencia prenatal, la respuesta estimada también ha sido baja y no diferente a la obtenida en los experimentos de selección directa por tamaño de camada, tanto en porcino (Johnson et al. 1999) como en ratón (Gion et al. 1990). No se ha llevado a cabo ningún experimento de selección por índices en conejo.

La selección por capacidad uterina tampoco ha tenido más éxito en la mejora del tamaño de camada que la selección directa en conejo (Blasco et al., 2005; Bolet et al., 1994; Mocé et al., 2005; Santacreu et al., 2005), porcino (Freking et al., 2016; Leymaster y Christenson, 2000) y ratón (Clutter et al., 1994; Kirby y Nielsen, 1993).

En conejo, se han realizado dos experimentos de selección divergente por capacidad uterina con dos criterios de selección diferentes. El criterio de selección en el INRA de Francia fue incrementar y disminuir el número de fetos muertos desde la implantación hasta el parto durante la segunda gestación en las hembras ULO. Este criterio solo consideraba la capacidad uterina en el periodo fetal. Después de tres generaciones de selección, no hubo respuesta a la selección para el número de fetos muertos, ni tampoco para la capacidad uterina (Bolet et al., 1994). En la Universidad Politécnica de Valencia en 1991 se inició una selección divergente para capacidad uterina y el criterio de selección fue el tamaño de camada de las hembras ULO. Después de diez generaciones de selección, la diferencia entre las líneas de alta y baja capacidad uterina fue de 1.5 gazapos para la capacidad uterina en hembras ULO y de 2.35 gazapos para la respuesta correlacionada para el tamaño de camada en hembras intactas, aunque ambas líneas mostraban una tasa de ovulación similar (Blasco et al., 2005; Santacreu et al., 2005). La diferencia entre las líneas alta y baja en tamaño de camada se debió a una mayor respuesta en

la línea baja. La respuesta estimada con una población control fue asimétrica. La disminución de tamaño de camada en la línea baja se debió a un descenso de la supervivencia embrionaria y de la fetal. Así, la línea seleccionada para aumentar la capacidad uterina, e indirectamente la supervivencia prenatal, no tuvo más éxito que la selección directa por tamaño de camada (Santacreu et al., 2005). Gran parte de las diferencias entre las líneas de alta y baja capacidad uterina, más del 50%, se produjeron en las dos primeras generaciones (Blasco et al., 2005) y los resultados de un análisis de segregación sugirieron la existencia de un QTL segregando en la población base para la capacidad uterina y el número de embriones implantados (Argente et al., 2003b).

#### **4. Genómica de los caracteres reproductivos**

Con el avance de la biología molecular y la genómica ha sido posible identificar las secuencias del DNA que explican una proporción significativa de la varianza de un carácter. Para los caracteres de producción de alta heredabilidad e importancia económica significó la validación de las genealogías (pruebas de paternidad) y la mejor predicción y progreso genético en los esquemas de selección. Por ejemplo, en el cerdo el descubrimiento de genes como el gen del halotano y QTLs de calidad de carne generó progresos importantes en la industria porcina (Oliver et al., 1993).

En el caso de los caracteres de naturaleza poligénica, como es el caso de los caracteres reproductivos, se requiere de herramientas genómicas muy potentes para poder identificar la estructura genética detrás de cada carácter. Para hacer uso de esa información es necesario conocer marcadores genéticos y mapas genómicos para la especie de interés. El tamaño de camada de una hembra, a lo largo de su vida reproductiva, es un carácter de una gran complejidad debido a que es el resultado de una sucesión de fenómenos que van desde el momento en que la hembra alcanza la pubertad hasta la consecución de varias gestaciones a lo largo de su vida (Vallet et al., 2010). Los genes influirían de forma secuencial en los resultados, en la medida en que si los primeros genes expresados tienen una actuación pobre o negativa, esto provocaría que la expresión de los posteriores genes en el desarrollo de la gestación pueda ser prácticamente indetectable, lo cual exige una correcta recolección de los datos fenotípicos (Stram, 2014; Vallet et al., 2010). Por el otro lado, el “conceptus” contribuye a la implantación, el desarrollo placentario, la supervivencia fetal durante la preñez, la supervivencia perinatal y el crecimiento desde el nacimiento al destete. Esto nos indica, que dentro de este sistema, los genes y la expresión de dos genomas (madre e hijo) podrían interactuar (Mocé et al., 2004;

Vallet et al., 2010). No obstante, la mayoría de los análisis de asociación genómica se realizan sobre las diferencias entre las hembras y se ignora la influencia de los genes de la cría, porque cada cría hereda la mitad de los genes de su madre (Vallet et al., 2010).

Por tal motivo, es necesario utilizar las ciencias genómicas y sus herramientas, en conjunto con herramientas estadísticas para encontrar resultados coherentes y válidos de asociación entre el genoma de la especie y su expresión fenotípica (carácter productivo).

#### **4.1. Frecuencia Alélica y Deriva Génica**

Los alelos de un gen cuya frecuencia sea mayor del 1% en un genoma de referencia y en toda una población, se denominan polimorfismos. Estos polimorfismos pueden ser por tamaño (microsatélites) o por sustitución de variantes (SNP) (Stram, 2014). En el genoma una variación en una posición nucleotídica se conoce como SNV (Variante de Nucleótido Simple), cuando esta variante se presenta con una frecuencia mayor al 1% se denomina SNP (Polimorfismo de Nucleótido Simple). La distinción entre SNV y SNP es arbitraria ya que la frecuencia del SNV es generalmente dependiente de la población de referencia (Cockett, 2010).

Los SNPs son los polimorfismos más utilizados actualmente porque representan la variante más pequeña en una cadena del DNA, tienen naturaleza bivalente (porque generalmente solo admiten dos alelos en el mismo lugar) y pueden relacionarse con la varianza genética de un carácter. No obstante, se debe tener en cuenta que la mayoría de estos SNPs a lo largo del genoma tiene un efecto neutral sobre la expresión de los genes (Stram, 2014). Así pues, es necesario que nuestras poblaciones presenten polimorfismo para nuestro carácter de interés.

En poblaciones silvestres, las frecuencias alélicas de estos polimorfismos permanecen constantes por el equilibrio de Hardy-Weinberg cuando las fuerzas de la evolución permanecen constantes (Hedrick, 2011). En los programas de mejora genética se busca modificar este equilibrio a través de la selección con el fin de incrementar un rasgo deseado (más carne, más crías o más lana). Esta selección artificial modifica las frecuencias alélicas, favoreciendo aquellas que influyan en la presencia del carácter. No obstante, la mayoría de las poblaciones de selección artificial son de tamaño relativamente pequeñas, por lo cual la deriva génica puede tener un papel importante como disruptivo o ruido en los programas de mejora.

Por otro lado, uno de los principios fundamentales de la genética de poblaciones aplicable a todos los estudios de asociación genómica, es que por lo menos todos los portadores de una variante genética heredable dada provienen del más reciente ancestro común único (MRCA) (Hedrick, 2011; Stram, 2014; Wakeley, 2009). Utilizando el concepto de MRCA, el cambio de las frecuencias génicas puede ser modelizado utilizando la teoría de la coalescencia. La coalescencia simula los procesos de recombinación en una población para cada generación y permite modelar la historia de una población del presente al pasado. Para ello se asume que la probabilidad de coalescencia ( $P$ ) entre dos líneas o poblaciones en la anterior generación es uno dividido dos veces el tamaño de la muestra ( $1/2N$ ) (Wakeley, 2009). Por lo tanto la probabilidad de coalescencia en la generación ( $t$ ) sería:

$$P(t) = (1 - P)^{(t-1)} * P \approx P * e^{-P * t}$$

En donde  $t$  es el número de generaciones y  $e$  es la constante de Napier. La fórmula se acerca a una distribución exponencial si  $P$  es pequeña y  $t$  es grande; en donde el tiempo se calcula como:

$$E(t) = \frac{1}{P} = 2N$$

La historia de una muestra de tamaño  $n$  individuos comprende  $n - 1$  eventos de coalescencia. Se parte desde el presente cuando la muestra es  $n$  y decrece el número de linajes de forma factorial, hasta que finalmente se llega a uno. El resultado será un árbol genealógico de líneas hasta el ancestro común (Figura 1). Por otra parte, los métodos “gen-dropping” permiten modelar la historia de una población del presente al futuro a partir de unas frecuencias génicas iniciales y una estructura de pedigrí determinada (Hickey y Gorjanc, 2013; Wakeley, 2009).

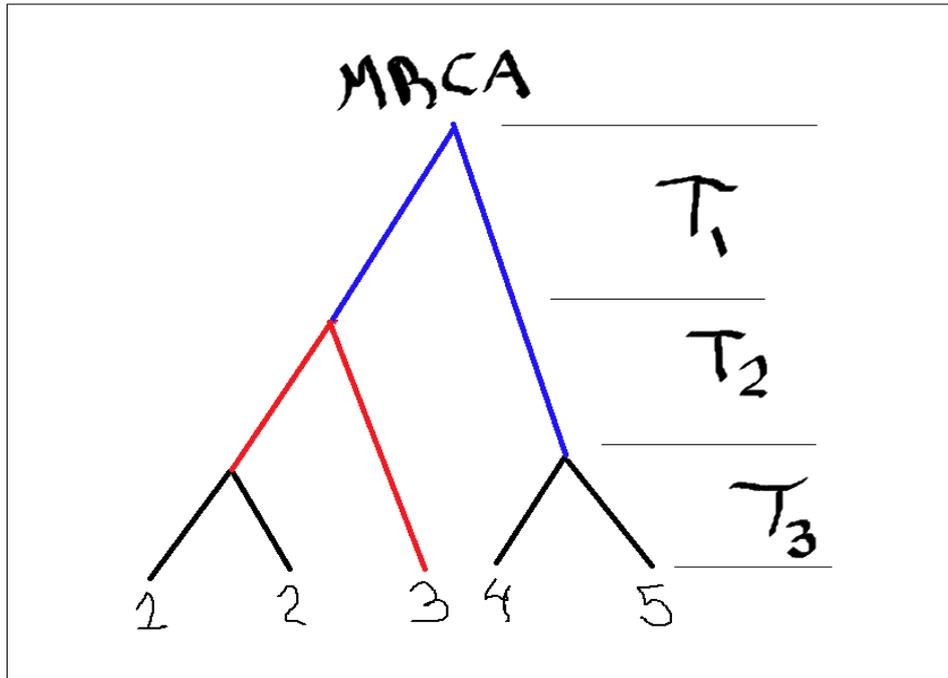


Figura 1. Árbol de una Coalescencia de una muestra de  $n= 5$  unidades

La teoría de la coalescencia y los métodos gen-dropping son ampliamente utilizados en los programas de simulación de poblaciones para simular procesos de selección genómica. Se puede modelar el desequilibrio de ligamiento, diferentes estructuras genéticas y diferentes procesos que modifican las frecuencias génicas, como por ejemplo la deriva génica (Hickey y Gorjanc, 2013).

La deriva génica tradicionalmente se ha medido a través del tamaño efectivo de la población ( $N_e$ ), que se define como la cantidad de individuos de una población que tiene la capacidad de reproducirse y dejar descendientes. Cuanto menor sea el tamaño efectivo de la población mayor probabilidad de influencia de la deriva (Falconer y Mackay, 1996). La cuestión está en que  $N_e$  solo es un indicador pero no una medida precisa de los efectos de la deriva. Al ser la deriva un proceso estocástico, se ha propuesto otra aproximación, como es por ejemplo determinar el tiempo en que un alelo podría fijarse en la población, a través de la ecuación de Kimura y Ohta (1971):

$$T(q) = - \frac{4N_e (1 - q) \ln(1 - q)}{q}$$

En donde  $T$  es el tiempo,  $q$  la frecuencia alélica de un alelo y  $N_e$  el tamaño efectivo de la población. La media del tiempo hasta la fijación es una función lineal del tamaño de la población y decrece cuando la frecuencia alélica inicial es grande, llegando a ser cercano a cero (Hedrick, 2011).

Si se quiere estimar el efecto de la deriva en las poblaciones simuladas hay que tener en cuenta que el resultado de la simulación será dependiente de la información a priori de la distribución de frecuencias alélicas iniciales (Hickey y Gorjanc, 2013; Stram, 2014). Cuando se utilizan marcadores de un panel de SNPs y poblaciones simuladas es posible probar la hipótesis sobre si el efecto del cambio de las frecuencias alélicas se debe solo a la deriva génica, o a una conjunción de la selección y la deriva en un proceso de selección artificial. Así por ejemplo, en un experimento de gallinas ponedoras seleccionadas divergentemente por comportamiento agresivo de picotazos, se pudo concluir que la selección tuvo mayor influencia que la deriva en el cambio de frecuencias génicas (Grams et al., 2015).

Todo esto nos ayudaría a resolver la pregunta sobre si la deriva génica está influyendo en poblaciones seleccionadas por tamaño de camada. Para esto tendríamos que realizar simulaciones con un gran número de SNP afectando al carácter, como sería para un carácter en el modelo infinitesimal. Esto serviría de modelo para el tamaño de camada en el conejo, de modo que las frecuencias alélicas dentro del genoma mostrarán polimorfismo o variabilidad para el progreso genético. Sin embargo, el conejo presenta pérdidas de variabilidad genética estimadas en promedio del 37.1% basado en el parámetro  $\pi$  y del 53.8% basado en el parámetro  $\theta$ , comparado entre la especie silvestre y la doméstica. Estas variables son estimadores de la variabilidad nucleotídica (Carneiro et al., 2011). Los estudios sugieren que el conejo comparado con otras especies ha perdido una gran diversidad genética durante el periodo de domesticación. La diferencia nucleotídica entre especies de conejos domésticos es de 0.188% a nivel autosomal y 0.205% de loci ligados al cromosoma X. Esto indica que hay poca variabilidad en las especies domesticas del conejo, lo cual es basado en la hipótesis de dos cuellos de botella (Carneiro et al., 2011, 2014a).

Si en el conejo se quisiera asociar su genoma con un carácter productivo es necesario tener polimorfismos y frecuencias alélicas no fijadas, pero se necesita además un estrategia o técnica con poder suficiente para identificar la localización de estos en el genoma. Con respecto a esto, todas las técnicas utilizadas intentan aprovechar la dependencia estadística entre los loci del genoma, término conocido como el desequilibrio de ligamiento (LD). Dos de estas técnicas se

tratarán en las dos secciones siguientes: genes candidatos y el barrido genómico o análisis de asociación de genoma completo (GWAS).

## 4.2. Genes candidatos para capacidad uterina en conejo

El método de genes candidatos requiere del conocimiento previo de la existencia del gen y su funcionalidad biológica. Los genes candidatos se eligen por su implicación fisiológica, pero del conjunto de genes candidatos unos genes contribuirán más que otros en la expresión del carácter. El interés se encuentra en detectar el gen o los genes que por pleiotropía tengan efectos sobre varios caracteres de la reproducción (Vallet et al. 2010; Cockett 2010). No obstante, la técnica comparada con otras técnicas (GWAS) tiene menor probabilidad de detección de genes que expliquen una parte importante de la variación genética de los caracteres reproductivos. Esto es debido a que se conoce poco acerca de los genes que intervienen en estos caracteres y el número de genes implicados en las diferentes rutas metabólicas (Vallet et al. 2010). No obstante, con esta metodología se han investigado algunos genes de interés en los caracteres reproductivos, como son por ejemplo, el receptor del estrógeno que está asociado en algunas poblaciones con el tamaño de camada en porcino (Rothschild et al., 1996), el gen de la hormona folículo estimulante beta (FSHB) y el gen de la osteopontina (OPN) (Spötter y Distl, 2006). Pero los efectos de estos genes dependen de la población animal estudiada.

En el pasado, los polimorfismos en un gen candidato se detectaban con el método de PCR-RFLP (Reacción de Cadena de la Polimerasa – Polimorfismo de Longitud de Fragmentos de Restricción) (Cockett, 2010). Esta metodología permite encontrar regiones del genoma de varios cientos de Kilobases (aproximadamente 20 cM), asociados con el carácter, a los cuales se les denomina loci de carácter cuantitativo (QTL). Los métodos de detección de QTLs presentan limitaciones, como por ejemplo, los grandes intervalos de confianza de los QTLs cuando los estudios presentaban bajo número de marcadores. Además, los tamaños poblacionales utilizados en especies animales de interés económico son relativamente bajos, lo cual limita la potencia de estos estudios de detección de QTLs (Ledur et al., 2010). Todo ello justifica que sólo en algunos casos se haya alcanzado identificar la mutación causal asociada a la segregación del QTL, que además tiene que ser validada genética y funcionalmente (Rothschild et al., 1996). Actualmente, el uso de esta técnica está disminuyendo por el uso de la técnica de barrido genómico, que puede utilizarse en el genoma entero siempre y cuando se tenga un genoma de referencia y un panel de SNPs.

Con respecto al conejo, se han realizado estudios de asociación en las líneas divergentes del experimento de selección por capacidad uterina de la UPV, descrito anteriormente. Con un análisis de muestra de una cruzamiento F<sub>2</sub>, se pudo concluir que la uteroglobina no presentaba asociación, pero el gen receptor de la progesterona y la oviductina presentaba alguna evidencia (Merchán et al., 2004). La expresión del gen glicoproteína oviductal 1 (OVGP1) presentó efectos positivos sobre el número total de gazapos, número de nacidos vivos, número de embriones implantados y supervivencia embrionaria y fetal. Sin embargo, falta la validación genética y funcional en otras poblaciones, porque el polimorfismo estudiado puede estar en desequilibrio de ligamiento con la mutación causal (García et al., 2010; Merchán et al., 2009).

También, por análisis de expresión usando las técnicas de “hibridación sustractiva supresora” (SSH) y PCR cuantitativa con retrotranscriptasa (RT-qPCR) se identificaron que los transcritos de los genes PGR, HSD17B4 y EROL estaban sobre expresados en tejido de oviducto recuperado a las 62 horas después del apareamiento en la línea baja (Ballester et al., 2013)

En otro experimento con animales de la séptima generación relajada, posterior a las diez generaciones de selección por capacidad uterina, se realizaron otros estudios por genes candidatos (IGF1 mRNA, el gen SGB1A1 y el gen “tejido inhibidor de la metaloproteinasas 1” - TIMP1). Ninguno de estos genes tuvo relevancia para el tamaño de camada. Sin embargo, TIMP1 mostró cierta asociación con la supervivencia embrionaria (Argente et al., 2010)

En cuanto al gen del receptor de la progesterona se encontró un SNP en la región promotora (g.2464G>A). El alelo G fue encontrado en un 75% de la línea alta y 29% en la línea baja. El genotipo GG presentó efectos de 0.5 gazapos y 0.5 embriones implantados más que el genotipo AA a las 72 horas. GG presentó 0.36 embriones más que AA e igualmente más avanzado el estado de desarrollo embrionario, mostrando un porcentaje más alto de mórula compactada. Entre el heterocigoto y el homocigoto GG se encontraron diferencias para tamaño de camada y número de nacidos vivos (Peiró et al., 2008). Un análisis posterior por “Western Blot”, demostró diferentes concentraciones de las isomorfias del receptor de la progesterona entre las líneas alta y baja, atribuidas al genotipo más frecuente (GG) (Peiro et al., 2010).

Los indicios de ciertos genes con efectos de pleiotropía que mostraron supuestas asociaciones causales, así como la alta respuesta a la selección en las dos primeras generaciones de capacidad

uterina en conejo, fomenta el interés de analizar estos resultados a través de técnicas más potentes como los barridos genómicos a lo largo de todo el genoma.

### 4.3. Estudio de asociación del genoma completo (GWAS)

El GWAS se basa en la presunción de que un alelo con efecto moderadamente grande ha segregado de uno o pocos individuos ancestrales y que los haplotipos con este alelo han pasado a las siguientes generaciones por selección natural o artificial, proceso conocido como barrido selectivo (“selective sweep”) (Stram, 2014). El haplotipo relacionado con el carácter en la población actual sería identificado fácilmente si se contara con todos los haplotipos salvajes. Esta herramienta estadística se apoya en una técnica de PCR alelo-específica usando oligonucleótido alelo-específicos (ASOs). Los SNP Arrays (SNPchips) son extensiones de los métodos ASOs dentro de una membrana o “bead”, por la cual múltiples alelos o marcadores genéticos pueden ser analizados en una simple ejecución y de manera conjunta (Cockett, 2010).

Previo al análisis de GWAS, un paso indispensable para el éxito del análisis es el desarrollo de los procesos de control de calidad de los datos genómicos (Turner et al., 2011). El procedimiento consiste en eliminar muestras de DNA (con respuestas dudosas por calidad de DNA o de procesamiento) y en la eliminación de marcadores del experimento con frecuencias demasiadas extremas para un locus (menor de 1%). Su objetivo principal es evitar errores tipo I (falsos positivos) y errores tipo II (falsos negativos) generados por los datos (Anderson et al., 2010; Turner et al., 2011). Entre los métodos de control de calidad existen postulados importantes como: (a) quitar un marcador genético es potencialmente de mayor impacto que quitar un individuo. (b) Un óptimo control de calidad debe quitar un pequeño porcentaje de muestras y SNPs sin disminuir la potencia total del estudio de asociación (Anderson et al., 2010) (c) y el umbral para eliminar por la menor frecuencia alélica (MAF) depende de la cantidad de muestras disponibles. Para pequeñas cantidades de muestras es necesario quitar SNPs con una frecuencia menor a 0.01, en cambio para gran cantidad de muestras estos SNPs raros serán necesarios para el experimento porque estarán a frecuencias muy bajas (Turner et al., 2011)

Para el GWAS se utilizan mapas genómicos o físicos, gran cantidad de muestras para la abundante cantidad de parámetros a estimar (efectos de SNPs) y es necesario un modelo estadístico apropiado a fin de explicar los efectos aditivos, dominantes y epistáticos (Ball, 2013;

Cockett, 2010; Stram, 2014). La técnica se basa en localizar SNPs asociados a la variante causal (polimorfismo) dentro de bloques de desequilibrio de ligamiento. Este desequilibrio de ligamiento (LD) puede medirse calculando las frecuencias observadas y esperadas de los haplotipos, para entonces calcular los parámetros estadísticos  $D'$  (estandarización del LD) y  $r^2$  (coeficiente de correlación de Pearson) (Du et al. 2007):

$$D_{ij} = P_{ij} - p_i q_j$$

$$D' = \frac{\sum_{i=1}^2 \sum_{j=1}^2 p_i q_j \frac{|D_{ij}|}{D_{max}}}{\sum_{i=1}^2 \sum_{j=1}^2 p_i q_j}$$

$$r^2 = \frac{D_{ij}}{p_1 p_2 q_1 q_2}$$

En donde

$$D_{max} = \begin{cases} \min [p_i q_j, (1 - p_i)(1 - q_j)] & \text{if } D_{ij} < 0 \\ \min [p_i (1 - q_j), (1 - p_i) q_j] & \text{if } D_{ij} > 0 \end{cases}$$

Los parámetros  $D'$  y  $r^2$  oscilan entre 0 y 1. Los subíndices “i” – “j” representan los dos sitios, loci o haplotipos comparados; y “p” - “q” las respectivas frecuencias alélicas dentro de ellos. De este modo, se puede descubrir los potenciales polimorfismos causales de la varianza del fenotipo. Empero para caracteres cuantitativos complejos se requieren más estudios en las regiones donde se haya encontrado asociación (Ball, 2013; Cockett, 2010).

Para el análisis de asociación genómica no se requiere que los individuos muestrales estén emparentados y tampoco se necesitan genealogías, propiedad indispensable en los análisis por ligamiento. Sin embargo, la genealogía puede ser utilizada para el control de calidad. El

objetivo es examinar de forma independiente la asociación de cada polimorfismo nucleotídico con el fenotipo, asumiendo siempre como verdadera la asociación entre el carácter y el marcador SNP por el desequilibrio de ligamiento (Bush y Moore, 2012). Para caracteres cuantitativos, se analiza la asociación de cada SNP con el fenotipo mediante modelos estadísticos de regresión lineal o métodos bayesianos, mientras que para fenotipos categóricos se utilizan tablas de contingencia, regresión logística o test de Chi cuadrado (Hayes, 2013).

En paneles de alta densidad es necesario una gran potencia de detección del estadístico para que tenga suficiente robustez para detectar los efectos genómicos y no generar asociaciones espurias. Pueden utilizarse métodos frecuentistas, cuyo procedimiento para tomar una decisión no es correcto porque asume independencia entre cada SNP, y además están basados en el cálculo del p-valor del experimento, el cual no es una medida de la incertidumbre porque cambiaría su valor cada vez que se repitiera el experimento. Como alternativa se pueden utilizar métodos bayesianos, que permiten calcular la probabilidad de la muestra para el estado de asociación, pero esta utiliza una información subjetiva de la distribución de los SNPs que puede tener relevancia en los resultados (Hayes, 2013). Para garantizar que la respuesta no es al azar, en los enfoques bayesianos para la genómica se utilizan fuertes evidencias o factores de Bayes altos para obtener mayor probabilidad de diferencias estadísticas. Por lo tanto, antes de comenzar a analizar datos, previamente se debe tomar en cuenta los factores que afectan la potencia del estadístico para detectar las asociaciones con los caracteres cuantitativos. Entre estos factores tenemos: las frecuencias alélicas del marcador y el QTL, la magnitud del desequilibrio de ligamiento, la heredabilidad del QTL, una prueba del poder de detección y el modelo genético (dominante, recesivo o aditivo) (Ball, 2013). No obstante, los antecedentes de posibles QTLs del carácter tamaño camada son escasos en el caso del conejo.

Por otro lado, para un GWAS de regresión lineal simple de SNPs, el escoger un nivel de significación es difícil porque se predispone a la presencia de falsos positivos. Por lo cual, la única evidencia de que una asociación detectada por GWAS es real es la validación en una población independiente. Para un análisis frecuentista en que la probabilidad de hallazgo tendría un error de  $P < 0.05$  se tendría un falso positivo por cada 20 evaluados, presentándose el problema de los test de comparación múltiple (Vallet et al., 2010). Un modelo clásico para GWAS simple sería:

$$\mathbf{y} = \mathbf{W}\mathbf{b} + \mathbf{X}\mathbf{g} + \mathbf{e}$$

En donde la  $\mathbf{y}$  representa un vector de valores fenotípicos,  $\mathbf{W}$  es la matriz de incidencias de los efectos fijos,  $\mathbf{b}$  es un vector de los efectos fijos,  $\mathbf{X}$  es la matriz con el número de alelos del alelo de referencia del SNP (genotipado),  $\mathbf{g}$  es el efecto del SNP y  $\mathbf{e}$  es un vector con los efectos residuales (Hayes, 2013; Stram, 2014).

Los modelos de métodos bayesianos de múltiple regresión tienen la ventaja que la potencia de detección no está inversamente relacionada con el número de marcadores (como en los métodos frecuentistas). Para calcular la estimas del efecto de los SNPs en los métodos bayesianos pueden utilizarse el Bayes A, B, Lasso, C o  $C\pi$  (Fernando y Garrick, 2013). A pesar que con estos métodos podría garantizarse una potencia alta, esto no parece resolver el problema principal del GWAS. El principal problema asociado a los resultados del GWAS es que la varianza genómica de los efectos de los SNPs no explica la totalidad de la varianza genética aditiva de los caracteres complejos, término denominado “heredabilidad pérdida”. Para resolver este problema se ha sugerido la utilización de poblaciones muestrales con tamaños suficientemente grandes, que superen el número de parámetros estimados, para poder tener mayor potencia y así poder obtener mejores estimaciones de la variación genética de los SNPs (Manolio et al., 2009). Esta situación podría ayudar en materia de estudios humanos, pero esta estrategia estaría limitada en las poblaciones de animales.

Además de los análisis de asociación entre SNP y el fenotipo, también es posible identificar áreas genómicas seleccionadas a través de índices de diferenciación genética ( $F_{st}$ ) (Johansson et al., 2010). Por ejemplo, cuanto mayor sea el número de generaciones de selección divergente que se lleven a cabo para un mismo carácter, mayor posibilidad se tendrá de identificar alguna región genómica, utilizando un número reducido de individuos de las dos poblaciones divergentes. En el caso de un experimento en líneas divergentes seleccionadas por comportamiento agresivo de picotazos en gallinas ponedoras, se encontraron regiones genómicas con grandes  $F_{st}$ . Este carácter presentaba baja heredabilidad en la línea alta (0.15) y también en la línea la baja (0.01) (Grams et al., 2015). Con respecto al conejo, los niveles de diferenciación genética ( $F_{st}$ ) son de 17.9% entre la especie silvestre y las razas cunícolas domesticadas (Carneiro et al., 2011).

Hoy en día se cuenta con diversas bases de datos con la información resultante de todos los análisis de detección de polimorfismos genómicos asociados con fenotipos de interés. En el ratón se cuenta con: Mouse QTL Data (<http://www.informatics.jax.org/mgihome/projects/overview.shtml>), mientras que para el cerdo se tiene: Pig QTL Data (<http://www.animalgenome.org/cgi-bin/QTLdb/SS/index>). En el caso del conejo, que es la especie de nuestro interés, sólo se cuenta con algunos QTLs de producción y calidad de carne no agrupados en una base de datos común (Sternstein et al., 2015).

Con información de estas y otras bases de datos de QTLs podemos mencionar el éxito de la técnica del GWAS en especies productivas para caracteres reproductivos. Por ejemplo: en ovino, Demars et al. (2013) identificaron dos mutaciones en el gen *BMP15*, en una región muy conservada del cromosoma X, vinculadas a una alta prolificidad. En cerdas se ha realizado estudios a través de aproximaciones bayesianas donde se ha descubierto QTLs para caracteres como lechones nacidos muertos y el intervalo entre partos. Estos estudios presentaron genes candidatos que están afectando a los dos caracteres: como el gen *ADAMTS9* y el gen *COL19A1*. Ellos también encontraron genes que sólo estaban asociados al intervalo entre partos, como el gen *TBC1D1* y el gen *SEMA4C*. Además, presentaron otros genes candidatos que están afectando solamente a los nacidos muertos *NUBPL*, *LOC100518697* y *DPP10* (Schneider et al., 2015). Para la tasa de ovulación en porcino, algunos investigadores encontraron un gran número de QTLs asociados al carácter. Del total de QTLs 22 de ellos explicaban el 71.10% de la varianza genética de la tasa de ovulación, incluyendo en estas regiones los genes del Receptor del Estrógeno 1, del Factor de Diferenciación de Crecimiento 9, y de la Inhibina  $\beta$ A. Esta información podría ser utilizada en los esquemas de mejora genética de cerdo y en las poblaciones comerciales de cerdos (Schneider et al., 2014). Además, otros investigadores han detectado varias regiones genómicas asociadas con caracteres de longevidad reproductiva en porcino, lo que indicaba la presencia de una naturaleza poligénica con un gran número de QTLs presentes (Onteru et al., 2011b).

Estos resultados de asociación en especies domésticas son debido a los bloques de desequilibrio de ligamiento en las poblaciones estudiadas. En conejo, los bloques de desequilibrio de ligamiento entre razas son de menor tamaño que dentro de razas. La extensión de estos bloques dentro de razas puede llegar de 0.40 a 3.20 Mb con  $r^2$  mayores de 0.2. Estos valores indican una estructura de bloques de desequilibrio de ligamiento grande, lo cual representaría un beneficio para la reducción del intervalo de asociaciones en futuros estudios de mapeo genético. En otras palabras, una escala más precisa para asignar las posiciones de algunos

SNPs y la determinación de extensiones de LD dentro de razas, ya que estructuras definidas en bloque sobre el genoma implicarían mejor potencia de detección de caracteres complejos, utilizando poblaciones divergentes seleccionadas para experimentos con cruzamientos (Carneiro et al., 2011).

Actualmente se cuenta con una base de datos del genoma de referencia del conejo denominado OryCun2.0 desarrollado por Secuenciación con Sanger, un “Conting” N50 de 64.7 Kb y un “scaffold” N50 de 35.9 Mb de tamaño, realizando un tamaño de ensamblaje total de 2.55 Gb. Además se ha puesto a disposición de la investigación el primer SNP-Array para conejos (Carneiro et al., 2014b). Por consiguiente, la disponibilidad de esta herramienta genómica y de muestras únicas de individuos proveniente de una selección divergente por capacidad uterina, presenta la opción de realizar una investigación para buscar genes asociados al tamaño de camada, teniendo una mayor posibilidad de encontrar asociaciones relevantes porque el objetivo de la selección fue por un único carácter (la capacidad uterina).

## OBJETIVOS

La estructura genética de los caracteres reproductivos del conejo de producción de carne es desconocida. El objetivo es elucidar algunos conocimientos sobre este tema a partir de datos de un experimento de selección divergente por capacidad uterina durante diez generaciones y una población de referencia en conejo, mediante:

A. Estudio de simulación del experimento de selección divergente por capacidad uterina para:

- Evaluar la capacidad del GWAS para detectar asociaciones entre SNPs y áreas genómicas causales en diferentes escenarios de estructura genética.
- Estimar el efecto de la deriva génica en la selección divergente de capacidad uterina.

B. Análisis por GWAS y huellas de selección a partir de muestras de DNA de las líneas divergentes y una población de referencia para:

- Determinar la presencia de áreas genómicas asociadas al tamaño de camada y sus componentes.
- Determinar los efectos de la selección por capacidad uterina sobre el genoma en las líneas divergentes y la población de referencia.
- Identificar posibles genes candidatos por posición debido a la selección.

# ARTICULO 1: ESTRUCTURA GENÉTICA DE DOS LÍNEAS DIVERGENTES SELECCIONADAS PARA CAPACIDAD UTERINA EN CONEJO.

## Resumen

**Antecedentes:** En los caracteres complejos la naturaleza genética es desconocida. Para evaluar los diseños experimentales de estudios de asociación del genoma completo (GWAS) se hace necesario analizar diferentes escenarios donde el número de regiones genómicas causales disponga de resultados similares a las poblaciones reales. El objetivo fue determinar el diseño óptimo de estructura genética para la detección de asociaciones positivas en poblaciones simuladas, utilizando información genealógica de dos líneas de un experimento de selección divergente por el carácter capacidad uterina en conejas que previamente había sido seleccionada por tamaño de camada al destete.

**Métodos:** Se evaluaron dos estrategias con el mismo número total de animales, una en la que solo se disponía de una línea única seleccionada por capacidad uterina y otra en la que se disponía de dos líneas divergentes seleccionadas por capacidad uterina. Tres grupos con distintos números de QTNs (10010, 1010 y 110) se simularon para una población que primero se seleccionó a favor de tamaño de camada al destete por 12 generaciones, para luego seleccionarse divergentemente por capacidad uterina durante 10 generaciones. Se evaluaron bajo un método de regresión simple por marcador (SMR), clasificando como asociación verdadera de SNP, aquellos a una distancia  $\leq 1$  Mb y un  $r^2 \geq 0.5$  con respecto al QTN más cercano. Se utilizó un umbral calculado por una corrección de test múltiple por Bonferroni para determinar las asociaciones positivas o negativas. Con las asociaciones clasificadas se analizaron el error tipo I (T1E), la potencia de detección y el valor predictivo positivo (PPV). Además, se evaluó el efecto de la deriva utilizando la respuesta a la selección y el porcentaje de QTNs fijados.

**Resultados:** La simulación mostró que utilizando 160 animales de la línea única, la regresión simple por marcador (SMR) no fue capaz de determinar alguna asociación positiva por encima del umbral. En cambio, utilizando individuos de las líneas divergentes se encontraron asociaciones positivas. El porcentaje de error tipo I fue elevado (37, 48, 36%) y el análisis de potencia fue alto para todos los grupos (93, 97, 83%). Por lo tanto, no hubo diferencias relevantes en estas pruebas, no obstante, el valor predictivo positivo fue mejor para el grupo con 10,010 QTN (57%) frente a 1,010 (29%) y 110 QTN (8%). Con respecto a la deriva, su

efecto en la selección llegó a ser del 35 al 50% de los QTNs fijados, y de 10.2 a 14.5% de la respuesta a la selección por capacidad uterina.

**Conclusiones:** Para estudios de asociación del genoma completo (GWAS), la ventaja de poseer líneas divergentes aumenta la posibilidad de encontrar asociaciones positivas. El grupo de 10,010 QTN tiene un mayor valor predictivo positivo, pero un alto valor de error tipo I. Con esta estructura se obtendría mayor cantidad de asociaciones reales e implicaría un modelo infinitesimal por la cantidad de efectos pequeños que afectan a la capacidad uterina. No obstante, un análisis con datos reales bajo esta estructura y con líneas divergentes supondría la aparición de abundantes asociaciones a lo largo del genoma, pero produciendo poca discriminación entre los verdaderos y los falsos positivos. Además, se infiere que la deriva génica presentó poco efecto sobre la respuesta a la selección.

**Palabras Claves:** GWAS, error tipo I, análisis de potencia, valor predictivo positivo, capacidad uterina, deriva génica.

## Introducción

En la actualidad las tecnologías emergentes de la genómica hacen factible realizar estudios de asociación del genoma completo. Estos estudios identifican regiones genómicas que afectan fenotipos complejos de interés económico (Stram, 2014). No obstante, el desconocimiento de la estructura genética de estos caracteres conlleva una incertidumbre sobre el diseño de los experimentos y los métodos adecuados para ser más eficientes en la identificación de regiones genómicas de interés. Una alternativa para ayudar a entender cómo proceder en estas investigaciones son las simulaciones para comprobar la potencia de detección de áreas genómicas asociadas a determinados caracteres cuantitativos (Chen et al., 2009; Purcell et al., 2003).

En el Departamento de Ciencia Animal de la Universidad Politécnica de Valencia se cuenta con muestras de sangre de animales de dos líneas divergentes seleccionadas por capacidad uterina durante 10 generaciones. La capacidad uterina está relacionada con el tamaño de camada, que es uno de los caracteres económicamente más importantes en el conejo. Con el advenimiento del SNPchip de Affymetrix en conejo, surge el interés por detectar áreas genómicas asociadas con este carácter reproductivo. La detección de QTL con plataformas de genotipado para caracteres reproductivos de baja heredabilidad, se ha logrado con éxito tanto en ovino como en porcino (Demars et al., 2013; Onteru et al., 2011; Schneider et al., 2014).

El objetivo de nuestro estudio es estimar la detección de asociaciones positivas en diferentes grupos de estructura genética y para ello se realizan simulaciones, utilizando la información genealógica de las líneas implicadas en la selección. Además, otro objetivo es inferir el efecto de la deriva en la selección de capacidad uterina.

## **Material y Métodos**

### **Estudio de Simulación**

Esta investigación pretende simular poblaciones provenientes de dos procesos de selección. Primero se simuló 10 generaciones de apareamiento al azar previo al inicio de la selección para imitar la fundación de una línea denominada línea V. Después, el primer proceso de selección imitó una selección por tamaño de camada (NW) por doce generaciones de la línea V. A continuación, se inició un proceso de selección divergente por capacidad uterina (UC) durante diez generaciones. A la línea seleccionada a favor de la capacidad uterina fue denominada línea alta de UC y la línea seleccionada en contra de la capacidad uterina fue denominada línea baja de UC. La descripción de la selección de la línea V y la selección divergente de capacidad uterina se encuentra detallada en [Estany et al. 1989](#) y [Blasco et al. 2005](#), respectivamente.

Para la simulación se utilizó el programa AlphaSim 1.04 que combina los métodos de coalescencia y métodos “Gene Drop” ([Hickey y Gorjanc, 2013](#)). AlphaSim es un programa formado por un conjunto de programas. La simulación del programa consta de dos periodos. El primer periodo se realiza con el programa MaCS (“Markovian Coalescent Simulator”) para crear una población con el desequilibrio de ligamiento (LD). En este periodo, la asociación entre bases de la secuencia de los genomas es debida a la acción de la deriva y asume que los procesos de recombinación se ejecutan independientemente para lo que se utiliza una cadena de Markov ([Chen et al., 2009](#)). Para el segundo periodo de apareamiento al azar (burn-in) y selección, el programa AlphaSim se apoya en un conjunto de programas (AlphaMate, AlphaBayes, AlphaFormatter, AlphaAGH), de modo que puede ejecutar los apareamientos, las evaluaciones genéticas y la creación de datos genómicos y fenotípicos, entre otros procesos. Con el programa AlphaSim se consigue generar múltiples datos de genotipos y fenotipos con flexibilidad de acuerdo al establecimiento de estructuras de población (parentescos), la distribución de los efectos de nucleótidos asociados a un carácter cuantitativo (QTNs) y el panel de SNPs desde baja a alta densidad ([Hickey y Gorjanc, 2013](#)).

El programa MaCS simula muestras de haplotipos con la información de la secuencia para cada cromosoma de acuerdo a una población ancestral específica y tasas de recombinación y mutación fijadas por el usuario. El programa considera una tasa de recombinación de uno por cada 100 centimorgan (cM). Después de crear la población con desequilibrio de ligamiento son escogidos los SNPs que formaran parte del panel de SNPs, utilizando los genomas de los individuos de la última generación. De los sitios segregantes en el genoma que superen un valor mínimo de frecuencia son escogidos al azar los SNPs del panel. Los QTNs de los caracteres son escogidos del conjunto de todos los SNPs, los cuales pueden coincidir o no con los SNPs del panel. El programa permite definir valores mínimos y máximo de menor frecuencia alélica (MAF) de los QTNs en un rango de 0 a 0.05. Además el programa simula los efectos de los QTNs con una distribución normal o gamma.

El fenotipo de los animales para un carácter se genera por el modelo genético aditivo, definido como la suma del aporte genético más un residuo. La varianza residual permanece proporcional a la varianza de los valores mejorantes de los animales en la generación base del pedigrí, para asegurarse que la heredabilidad permanezca constante (Hickey y Gorjanc, 2013).

### Estructura del Conjunto de Datos Simulados

La simulación se estructuró en tres grupos de estructura genética en función al número de QTNs que determinan la parte heredable del carácter. El número de QTNs por cromosoma fue de 1000, 100 y 10 para los grupos A, B y C, respectivamente (Tabla 1). Además, se asignó un QTN por mutación para cada cromosoma con una heredabilidad de 0.0001, de este modo se diseña el efecto de la posible aparición de mutaciones durante la historia de la población. Los QTNs no están presentes en el panel de SNP y sus efectos son asignados con una distribución normal sin restricción para el rango de frecuencias de los QTNs.

Tabla 1. Número de QTNs por grupo de estructura genética.

Grupo	QTN / Cromosoma		QTN / Genoma		Total QTN
	Sitios Segregantes	Mutación	Sitios Segregantes	Mutación	
A	1,000	1	10,000	10	10,010
B	100	1	1,000	10	1,010
C	10	1	100	10	110

QTN: Nucleótido asociado a un carácter cuantitativo

Para cada grupo se hicieron 31 réplicas, haciendo un total de 93 simulaciones. Para crear el desequilibrio de ligamiento se utilizaron 2,000 haplotipos; diez cromosomas con un tamaño genómico de  $10^8$  pares de base por cromosoma, asumiendo una distancia genética de 100 centiMorgan (cM) y un tamaño de genoma de tres Gigabases, con una tasa de recombinación de  $3.33 * 10^{-9}$  y una tasa de mutación de  $2.5 * 10^{-8}$ . El tamaño efectivo de la población inicial ( $N_e$ ) fue de 100,000 animales y después de 1,000,000 generaciones la población se redujo a 500 animales. La tasa promedio de reducción de la población fue de 0.09994 animales por generación, con un incremento de la población en la última generación. Estos valores fueron asignados por razones de simplicidad y cómputo, pero de manera concordante y proporcional con el genoma del conejo y su domesticación (Carneiro et al., 2011).

Por cada replica de simulación se generaron 10 generaciones de apareamiento al azar (burn-in), 12 generaciones de la línea V seleccionada por el carácter tamaño de camada al destete (NW) y diez generaciones de las dos líneas del experimento de selección divergente por el carácter capacidad uterina (UC). Como se tienen registros genealógicos de todas las líneas, el número de la población total y de animales seleccionados por generación son variables, imitando a las poblaciones reales (Figura 1). La línea V presentó un  $N_e$  aproximado de 65 y las líneas alta y baja de UC presentaron 36 y 34, respectivamente. (Estos valores fueron calculados como una aproximación por la fórmula de  $N_e$  para números diferentes de padres y madres) (Falconer y Mackay, 1996).

Por otro lado, los SNPs fueron definidos como sitios segregantes con un MAF de 0.05 en la población base de la línea V. La población base de la línea V es la última generación de la población creada por el programa MaCS. Además, se confeccionó un SNP-Chip de 57,000 SNP (clasificado dentro del grupo de mediana densidad a razón de 5,700 SNP por cromosoma); de una manera aproximada y proporcional al número y tamaño promedio de cromosomas del conejo para el actual panel de Affymetrix de 200k SNP.

La selección se basó en un BLUP con información genealógica y sólo se utilizó el fenotipo de las hembras en todas las generaciones. La heredabilidad utilizada fue de 0.047 y 0.11, para NW y UC, respectivamente (Blasco et al., 2005; García y Baselga, 2002). La correlación genética y residual entre los dos caracteres fue de 0.739 y 0, respectivamente. El programa no permite seleccionar en contra de un carácter. Por lo tanto, para simular la línea baja de capacidad uterina se creó un carácter adicional con los mismos parámetros de la capacidad uterina, pero con una correlación entre ellos de -0.999. Para todos los caracteres, la varianza genética aditiva

utilizada antes del apareamiento al azar fue de 2 y una varianza residual proporcional a 20 y 8 veces la varianza genética aditiva para tamaño de camada al destete y capacidad uterina, respectivamente.

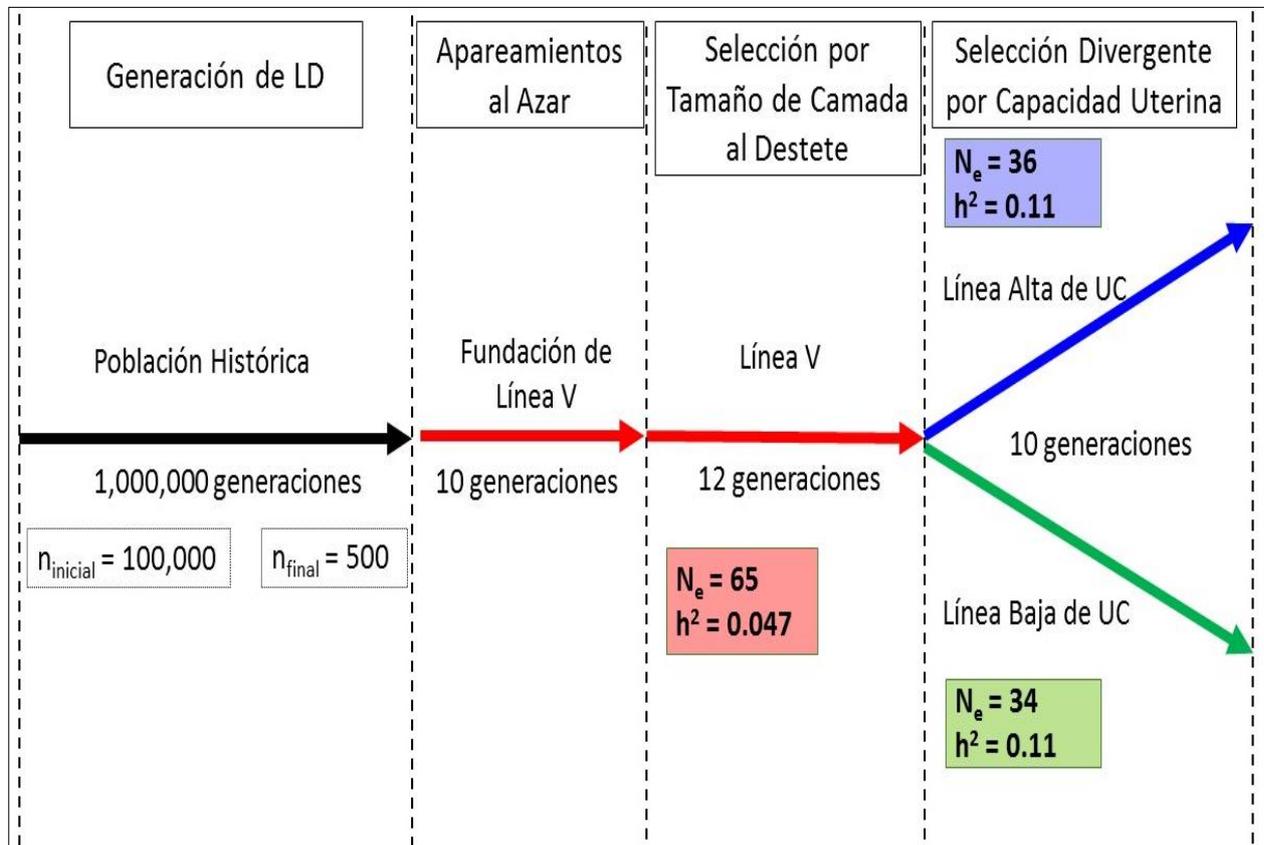


Figura 1. Esquema de simulación de la selección divergente de capacidad uterina.

### Comparación de Grupos de Estructura Genética

Para la última generación del experimento de selección divergente por capacidad uterina (generación 10) se seleccionaron 22 hembras en línea alta y 30 hembras en línea baja. Estas hembras produjeron 440 y 530 crías en la línea alta y baja, respectivamente. Después de este periodo, la selección fue relajada. Por lo tanto, para constituir la siguiente generación fueron elegidos los reproductores por un muestreo al azar. En total, los animales con fenotipo y genotipo fueron 90 y 70 animales de la línea alta y baja, respectivamente.

Para estudiar las ventajas de utilizar líneas divergentes versus sólo una línea seleccionada se evaluaron dos estrategias: (1) utilizando 160 animales de la línea alta (estrategia 1). Esta línea en este contexto se denominará como línea única. (2) Una estrategia utilizando 90 y 70

animales de líneas divergentes de la generación relajada (estrategia 2). Todos los pasos indicados en los siguientes párrafos fueron realizados en ambas estrategias.

Las pruebas de control de calidad se basaron en escoger SNPs cuyo MAF fuera mayor de 0.05. Además se determinaron los QTNs que estaban segregando en las poblaciones y se eliminó aquellos QTN fijados como se hace habitualmente.

Para llevar a cabo un estudio de asociación del genoma completo (GWAS) se utilizó un método de regresión simple para cada marcador (SMR), a fin de determinar las magnitudes de los p-valores de los alelos de cada SNP. Se calculó un umbral para cada simulación, utilizando un nivel de significación de  $P < 0.05$  ajustado al número de regresiones realizadas (test de Bonferroni) (Sham y Purcell, 2014). A continuación se muestra el modelo utilizado para el GWAS, que se ajusta independientemente para cada SNP:

$$\mathbf{y} = \mathbf{x}b_i + \mathbf{e}$$

En donde  $\mathbf{y}$  es un vector de dimensiones 160 x 1 con los valores fenotípicos de cada individuo y  $\mathbf{x}$  es un vector de dimensiones 160 x 1 con valores de 2, 0, ó 1 correspondientes al número de veces que aparece el alelo de referencia del SNP  $i$  en cada individuo. Esto significa que si el alelo A es el alelo de referencia, entonces el homocigoto AA = 2, el heterocigoto Aa = 1 y el homocigoto aa = 0, para todos los elementos del vector  $\mathbf{x}$ . El término  $b_i$  es el efecto del alelo de referencia del SNP  $i$  y  $\mathbf{e}$  es un vector de dimensiones 160 x 1 con los valores residuos de cada individuo. Los p-valores se calcularon con una prueba de  $X^2$ . Con los p-valores y el umbral de cada simulación se clasifican los valores predichos en: asociación positiva, si el valor de p-valor está por debajo del umbral, y asociación negativa, si el valor de p-valor está por encima del umbral (Figura 2a). Para considerar una asociación positiva como verdadera se consideraron dos condiciones utilizadas comúnmente: (a) Una distancia igual o inferior de una megabase entre el SNP y el QTN segregando más cercano ( $\pm 1$  Mb), (b) Una medida de desequilibrio de ligamiento ( $r^2$ ) igual o mayor de 0.5. Aunque la asociación entre un marcador genético y un gen causal no depende proporcionalmente de la distancia física, se asume que el desequilibrio de ligamiento (LD) disminuye conforme aumenta la distancia física por la acción de los eventos de recombinación en la historia de la población. Para el conejo en las razas domésticas se han encontrado bloques de desequilibrio de ligamiento (LD) desde 0.40 a 3.20 Mb con  $r^2 > 0.2$  (Carneiro et al., 2011); por consiguiente se estableció un condición mínima y

muy conservadora de  $r^2 \geq 0.5$  para que un SNP sea considerado como una asociación verdadera con el QTN.

Con estos criterios se determina la clasificación de los SNP, como se muestra en la Tabla 2. Las clases son: verdadero positivo (TP), falso positivo (FP), verdadero negativo (TN) y falso negativo (FN). Cuando estos valores son ordenados en una matriz de 2 x 2 se le denomina matriz de confusión (Figura 2b).

Tabla 2. Clasificación de SNP de acuerdo con la asociación

Tipo de SNP	Abrev. <sup>1</sup>	Descripción
Verdadero Positivo	TP	Distancia entre SNP y QTN más cercano $\leq 1$ Mb y $r^2 \geq 0.5$ p-valor $<$ umbral <sup>2</sup>
Falso Positivo	FP	Una distancia entre SNP y QTN más cercano $> 1$ Mb ó $r^2 < 0.5$ p-valor $<$ umbral <sup>2</sup>
Verdadero Negativo	TN	Una distancia entre SNP y QTN más cercano $> 1$ Mb ó $r^2 < 0.5$ p-valor $>$ umbral <sup>2</sup>
Falso Negativo	FN	Distancia entre SNP y QTN más cercano $\leq 1$ Mb y $r^2 \geq 0.5$ p-valor $>$ umbral <sup>2</sup>

<sup>1</sup> Abreviatura. <sup>2</sup> Umbral ajustado por el test de Bonferroni.

La evaluación de las dos estrategias (1 y 2) para los grupos de diferente número de QTNs, fue realizada a través del cálculo de tres parámetros (Balding, 2006; De Maturana et al., 2014; Purcell et al., 2003):

- a) Error Tipo I (T1E): es el cociente entre el número de falsos positivos y la suma de falsos positivos más verdaderos negativos, también llamado el ratio de falsas asociaciones positivas.

$$T1E = \frac{FP}{(FP + TN)}$$

- b) Análisis de Potencia: es el cociente entre el número de verdaderos positivos y la suma de verdaderos positivos más falsos negativos, también llamado el ratio de verdaderos positivos o razón de éxitos.

$$\text{Potencia} = \frac{\text{TP}}{(\text{TP} + \text{FN})}$$

- c) Valor Predictivo Positivo (PPV): es el cociente entre el número de verdaderos positivos y el total de positivos.

$$\text{PPV} = \frac{\text{TP}}{(\text{TP} + \text{FP})}$$

Además, se realizó un gráfico ROC (acrónimo en inglés de “Receiver Operating Characteristic”) para evaluar la coherencia de la regresión simple por marcador entre los tres grupos. Un gráfico ROC es una representación gráfica de la potencia frente al error tipo I para cada uno de los grupos evaluados. El punto en el gráfico es llamado un espacio ROC. En la Figura 2c se muestra un ejemplo de los espacios ROC donde la matriz de confusión puede cambiar sus valores de acuerdo al umbral utilizado ([Fawcett, 2006](#)).

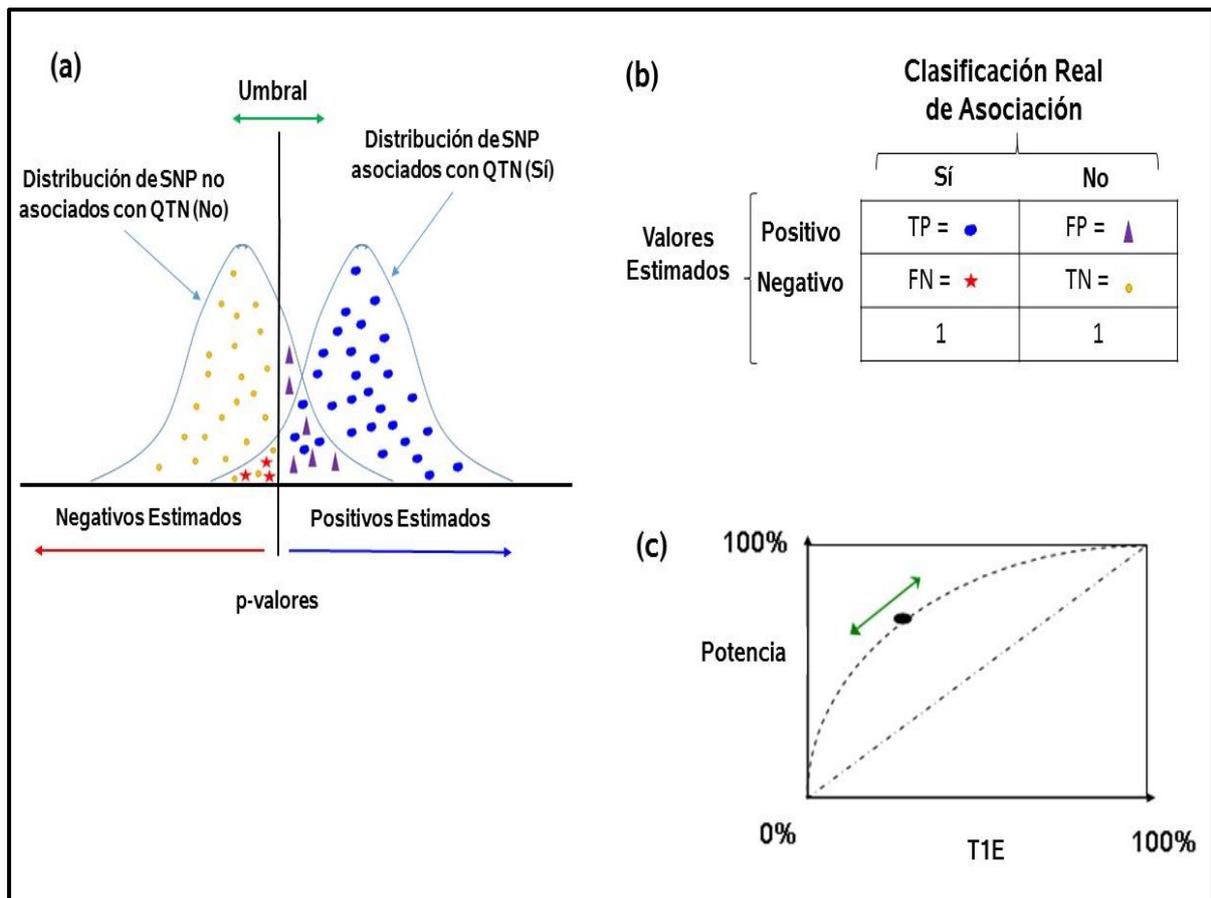


Figura 2. Ejemplo de creación de una matriz de confusión y un gráfico ROC. Para representar un gráfico ROC se siguen los siguientes pasos: (a) Determinación del umbral y la estimación de los p-valores de las asociaciones. (b) Creación de la matriz de confusión a partir de los p-valores y las condiciones para determinar si las asociaciones son reales o no. Estas condiciones son una distancia entre SNP y QTN más cercano  $\leq 1$  Mb y un desequilibrio de ligamiento ( $r^2$ )  $\geq 0.5$ . Las filas corresponden a los valores predichos o estimados por la regresión como positivo o negativo a partir de si los p-valores son mayores o menores que el umbral establecido mediante el test de Bonferroni. (c) Creación del gráfico ROC que relaciona la potencia de detección y el error tipo I (T1E). Un punto representa un espacio ROC. Si el punto está por arriba de la diagonal indica una mejor predicción, por abajo una peor predicción y exactamente en la diagonal equivale a un evento al azar.

### Efecto de Deriva Génica

Por otro lado, para determinar si el efecto de la deriva tuvo una gran influencia en la selección, se generaron con el grupo de 10,010 QTNs tres escenarios. En todos los escenarios se simuló las diez generaciones de apareamiento al azar antes de comenzar el proceso. Estos escenarios son: (A) escenario 1: 22 generaciones de apareamientos al azar, (B) escenario 2: 12 generaciones de selección en línea V más 10 generaciones de apareamientos al azar, y (C) escenario 3: 12 generaciones de selección en línea V más 10 generaciones de selección divergente de la línea UC. Para cada escenario se simuló 35 réplicas, haciendo un total de

105 simulaciones para el análisis. En los escenarios de deriva génica (escenario 1 y 2) para mantener el mismo número de tamaño efectivo de la población ( $N_e$ ) se utilizó en cada generación la misma proporción de hijos y apareamientos por progenitores, que la utilizada en el escenario de selección por UC (escenario 3).

## **Resultados**

El objetivo principal de este trabajo fue evaluar en qué medida se pueden detectar asociaciones entre los SNPs y la capacidad uterina, utilizando el mismo número de individuos de las líneas divergentes por capacidad uterina. Para ello se utilizó un modelo muy sensible al sesgo como la regresión simple por marcador (SMR) y además, una corrección conservadora de test-múltiple (test de Bonferroni) para la determinación del umbral.

Para evaluar los resultados de la selección de las poblaciones simuladas se observaron los valores de la varianza genética aditiva inicial y la respuesta a la selección de la capacidad uterina. Esto se evaluó debido a que el proceso de selección por tamaño de camada al destete modificó la varianza genética de la capacidad uterina. La respuesta a la selección es la suma de las respuestas obtenidas en las líneas alta y baja de capacidad uterina. Esta respuesta depende de la intensidad de selección, el intervalo generacional, la precisión de los valores mejorantes y la varianza aditiva inicial. Se esperaría que la respuesta fuera directamente proporcional a las diferentes magnitudes de la varianza aditiva inicial, ya que los primeros tres parámetros permanecen constantes en cada simulación. No obstante, en la Figura 3 se muestra como la respuesta final de la selección no tiene una alta correlación con la varianza aditiva inicial. Esto indica que procesos estocásticos (la deriva génica) podrían haber intervenido durante la selección. Este fenómeno será analizado posteriormente.

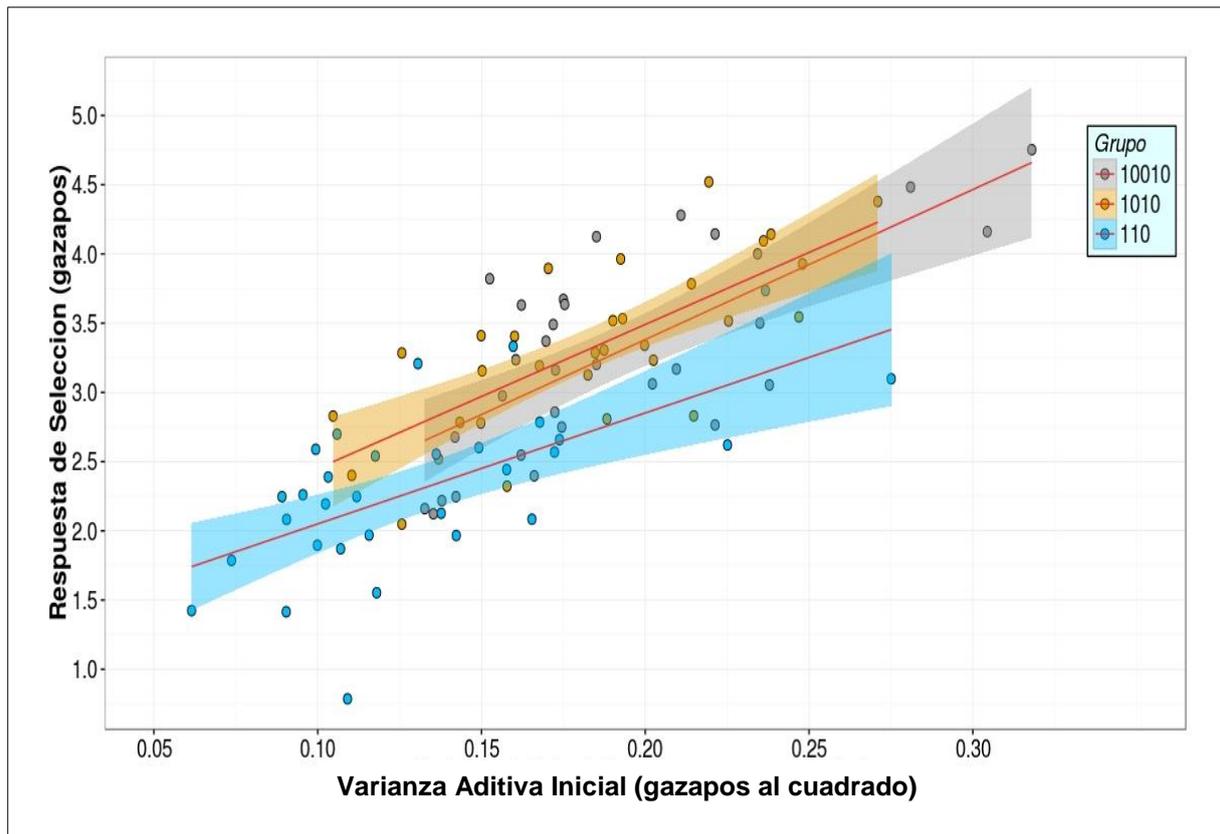


Figura 3. Relación entre la respuesta a la selección versus la varianza aditiva inicial. Cada punto en la gráfica corresponde a una población simulada. Las líneas continuas son las regresiones dentro de cada grupo. La sección sombreada corresponde a un intervalo de confianza de 95% para la regresión de cada grupo.

### Análisis por la Fijación de QTNs

Para el cálculo de asociaciones en cada grupo se determinó el número y porcentaje de QTN fijados tanto en la estrategia 1, como los fijados en la estrategia 2. En la Figura 4 se muestra los valores de QTNs fijados para cada uno de los tres grupos de QTNs. Las diferencias de QTNs entre la estrategia 1 y la estrategia 2 son sustancialmente importantes en términos de cantidad para el grupo de 10010 (Tabla 3). De acuerdo a los datos de la Tabla 3, la diferencia de QTNs entre utilizar la estrategia 1 o estrategia 2 estuvo alrededor de un 10% más de QTNs fijados en la línea única, para todos los grupos.

Una revisión más profunda de los valores de QTNs demostró que entre el 43 al 76% de los QTNs fijados en la estrategia 1 y no cuando se consideró la estrategia 2, tenían una magnitud de efecto superior al último cuartil de la distribución de efectos de QTNs. Por lo cual, se deduce que las 12 generaciones de selección por tamaño de camada, las posteriores 10 generación por

alta capacidad uterina y la alta correlación entre los caracteres, contribuyó a la fijación de estos QTNs en la línea alta de capacidad uterina. Cuando nosotros analizamos las frecuencias de las líneas divergentes (alta y baja) por separado de la estrategia 2, encontramos que estos QTNs fijados en la línea alta, presentaban en su mayoría frecuencias intermedias en la línea baja. Además, cuando se utilizó la estrategia 1, menos del 10% de los QTNs no fijados presentaban un alto MAF (mayor de 0.35), en contraste con la estrategia 2, en donde más del 70% de QTNs no fijados presentaban altos MAF (mayor de 0.35). Cuando la menor frecuencia alélica (MAF) es alta se incrementa la probabilidad de detección de un QTN; por consiguiente, en la estrategia 2 podrían detectarse más QTNs.

Tabla 3. Datos descriptivos de QTN-Fijados.

Grupo	Estrategia 1			Estrategia 2			Dif. QTN <sup>3</sup>	C.V. <sup>4</sup>
	Media (SD <sup>1</sup> )	Mediana	P <sup>2</sup> (%)	Media (SD <sup>1</sup> )	Mediana	P <sup>2</sup> (%)		
A	8,508.4 (243.3)	8,534	85.3	7,632.7 (315.7)	7,606	74.8	855	0.26
B	852.4 (35.7)	862	85.3	756.3 (40.1)	763	75.5	91	0.28
C	97.6 (5.0)	97	88.2	87.8 (4.8)	87	79.1	9	0.38

<sup>1</sup> SD: desviación estándar; <sup>2</sup> P: porcentaje de QTN-Fijados; <sup>3</sup> Dif. QTN: diferencias de QTN entre estrategia 1 y estrategia 2, calculada como la mediana de las diferencias de cada simulación dentro de grupo; <sup>4</sup> C.V.: coeficiente de variación de las diferencias de QTN entre la estrategia 1 y estrategia 2. La estrategia 1 es la utilización de una línea única y la estrategia 2 es la utilización de las dos líneas divergentes de capacidad uterina.

Por lo tanto el análisis de respuesta a la selección versus varianza aditiva inicial y las diferencias de QTNs fijados utilizando la línea única y líneas divergentes, demuestra que ocurrió un proceso de divergencia como resultado de la selección y la deriva.

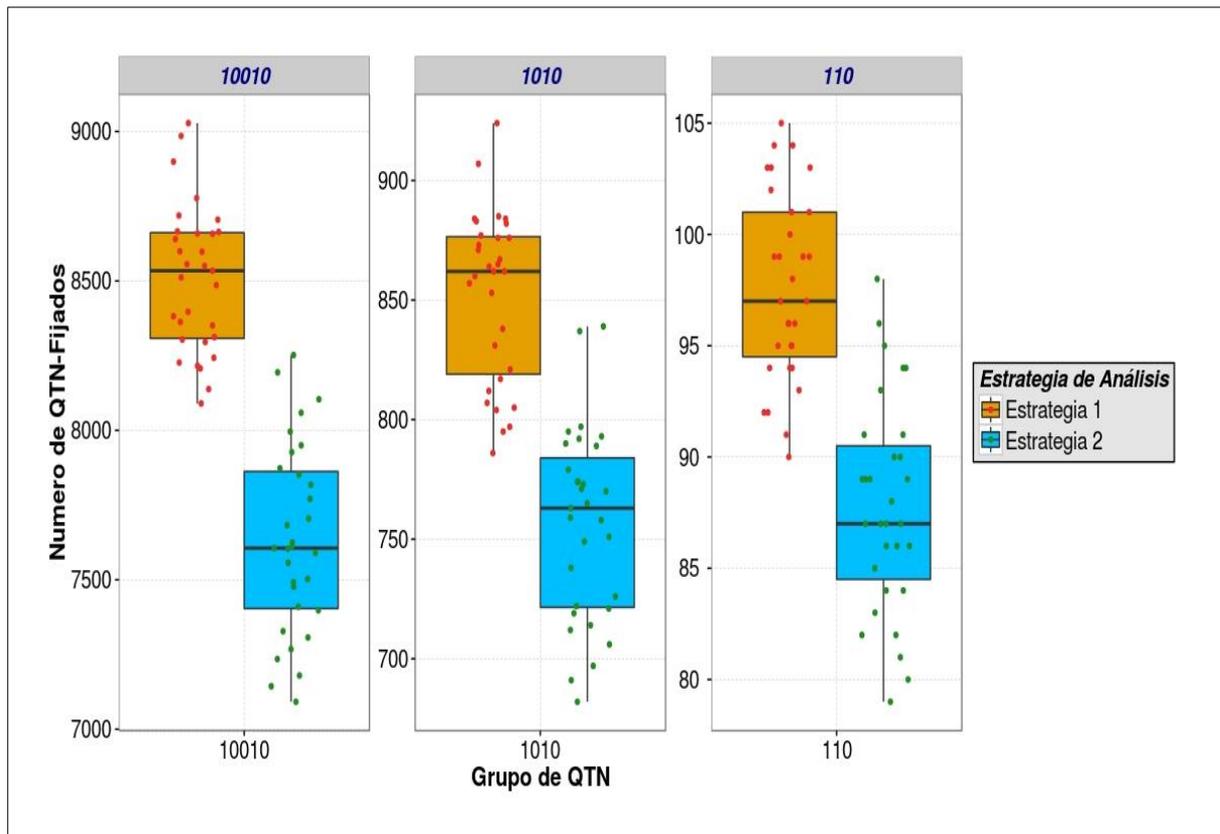


Figura 4. Número de QTNs fijados por cada grupo y estrategia. Los puntos rojos indican los valores para la estrategia 1 y los puntos verdes indican los valores para la estrategia 2. La estrategia 1 es la utilización de una línea única y la estrategia 2 es la utilización de las dos líneas divergentes de capacidad uterina.

### Comparación de Grupos de Estructura Genética

En cuanto a los análisis de la regresión simple por marcador (SMR) utilizando solo la estrategia 1, se encontró que en ninguno de los grupos de QTNs, la regresión pudo encontrar una asociación positiva en la línea única (estrategia 1), a pesar que existen SNPs que concordaban con los criterios de distancia  $\leq 1$  Mb y de desequilibrio de ligamiento  $r^2 \geq 0.5$  entre un SNP y un QTN.

En el caso de utilizar la estrategia 2 de líneas divergentes, la regresión sí encontró asociaciones positivas con la capacidad uterina. El resultado del número de QTNs segregando en la población, el número de SNP con un MAF mayor o igual a 0.05 y la matriz de confusión para cada grupo, se muestran en la Tabla 4. Estos resultados se calcularon como el promedio de los valores obtenidos de las simulaciones en cada grupo.

Tabla 4. Resumen de los resultados promedios por grupo utilizando la estrategia 2.

Grupo	Numero de QTN	Numero de QTN Segregando	Numero de SNP	Matriz de Confusión	
A	10,010	2,377	14,612	4,324 (TP) 240 (FN)	3,694 (FP) 6,324 (TN)
B	1,010	254	15,489	2,352 (TP) 63 (FN)	6,288 (FP) 6,786(TN)
C	110	22	15,382	392 (TP) 48 (FN)	5,228 (FP) 9,714 (TN)

TP: verdadero positivo, FP: falso positivo, TN: verdadero negativo, FN: falso negativo.

En la Tabla 4 se muestra que el número de QTN segregando en la población es proporcional al número de QTN. Además, el número de SNP que superan un MAF de 0.05 es independiente del número de QTN. En la matriz de confusión se muestra que la cantidad de falsos positivos es menor en el grupo A comparando con el grupo B y C. En cambio, el número de verdaderos positivos sí parece ser proporcional al número de QTNs.

El análisis del error tipo I (ratio de falsos positivos) muestra pocas diferencias entre las medias de los tres grupos de QTN estudiados (Tabla 5). Además, se muestra que el coeficiente de variación es mayor en el grupo C, y que la simulación que generó el máximo valor del error tipo I también pertenece al grupo C.

Tabla 5. Parámetros descriptivos del error tipo I (T1E) usando la estrategia 2 (líneas divergentes)

Grupo	Numero de QTN	Media	Mediana	SD <sup>1</sup>	Rango		CV <sup>4</sup>
					Min <sup>2</sup>	Max <sup>3</sup>	
A	10,010	36.7	38.9	16.5	0.2	62.1	0.45
B	1,010	47.7	48.9	14.1	11.9	68.8	0.29
C	110	35.5	37.3	20.0	0.0	73.0	0.56

<sup>1</sup> SD: Desviación Estándar, <sup>2</sup> Min: Mínimo; <sup>3</sup> Max: Máximo; <sup>4</sup> CV: Coeficiente de Variación

La Tabla 6 muestra los resultados descriptivos del análisis de la potencia de detección de SNP usando líneas divergentes. Los grupos presentan medias y medianas similares, sin embargo el

grupo C presenta un coeficiente de variación más elevado al igual que ocurría para el error tipo I.

Tabla 6. Parámetros descriptivos de la potencia de detección de asociación de SNP usando la estrategia 2 (líneas divergentes)

Grupo	Numero de QTN	Media	Mediana	SD <sup>1</sup>	Rango		CV <sup>4</sup>
					Min <sup>2</sup>	Max <sup>3</sup>	
A	10,010	93.4	98.4	17.5	4.6	100.0	0.19
B	1,010	97.2	100.0	6.2	71.0	100.0	0.06
C	110	83.2	100.0	31.5	0.0	100.0	0.38

<sup>1</sup> SD: Desviación Estándar, <sup>2</sup> Min: Mínimo; <sup>3</sup> Max: Máximo; <sup>4</sup> CV: Coeficiente de Variación

En la Tabla 7 se puede observar que las medias del valor predictivo positivo (PPV) son diferentes entre los grupos. Los resultados nos indican que podría predecirse mejor las asociaciones positivas verdaderas en un grupo con un mayor número de QTNs cuando se utiliza una regresión simple por marcador (SMR).

Tabla 7 Parámetros descriptivos del valor predictivo positivo usando la estrategia 2 (líneas divergentes)

Grupo	Numero de QTN	Media	Mediana	SD <sup>1</sup>	Rango		CV <sup>4</sup>
					Min <sup>2</sup>	Max <sup>3</sup>	
A	10,010	57.1	53.4	12.0	37.4	92.5	0.21
B	1,010	28.9	28.0	7.8	16.2	50.5	0.27
C	110	8.2	6.8	7.5	0.0	45.9	0.92

<sup>1</sup> SD: Desviación Estándar, <sup>2</sup> Min: Mínimo; <sup>3</sup> Max: Máximo; <sup>4</sup> CV: Coeficiente de Variación

Los resultados se resumen gráficamente en la Figura 5 mediante un “Violin-plot”. En este gráfico se observa que no existen diferencias entre las densidades de los grupos para los parámetros del error tipo I y de la potencia de detección. En cambio, para el valor predictivo positivo sí se observan diferencias en la densidad de probabilidad entre los grupos.

El conjunto de estos resultados indican que se obtiene el mismo error tipo I y la misma potencia de detección si el carácter de interés está determinado por un número de QTNs dentro del rango estudiado, entre 110 y 10010. Estos dos parámetros no parecen depender del

número de QTNs que determinan el carácter, sin embargo sí que se encontró una relación positiva entre el valor predictivo positivo y el número de QTNs estudiados. De los tres parámetros estudiados, la prueba del error tipo I es la más importante. Esto se debe a que un valor alto de error tipo I conlleva una alta proporción de falsos positivos, lo cual no permite discriminar cuales son las asociaciones verdaderas entre el total de las asociaciones positivas. Así por ejemplo, en una de las simulaciones del grupo de 10,010 QTN se encontraron las siguientes asociaciones: 3,883 TP, 1,988 FP, 5,724 TN y 271 FN. El error tipo I calculado a partir de estos datos es de un 26% y la potencia de detección es del 93%. De las 5,871 asociaciones positivas, que suponen el 93 % de todas las asociaciones en el genoma, es difícil determinar cuáles son las asociaciones verdaderas porque el 26% de los SNP no asociados se están presentando como asociaciones positivas.

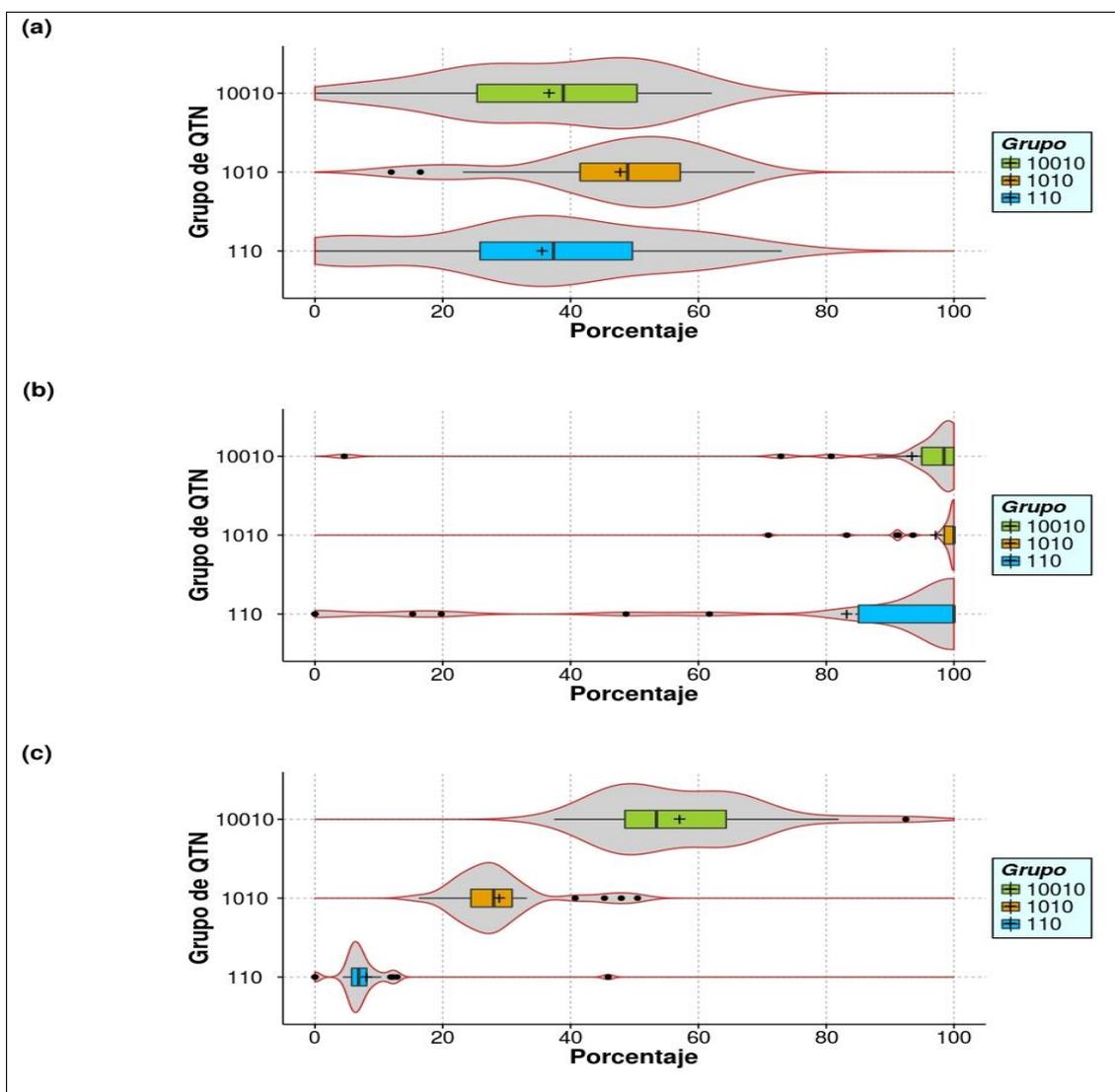


Figura 5. Violin-Plot. (a) error tipo I, (b) potencia de detección, (c) valor predictivo positivo. El signo de suma en color negro indica la media.

Por otro lado, en la Figura 6 se presenta el gráfico ROC con 93 espacios ROC que corresponden a las 93 simulaciones que se han llevado a cabo para la estrategia 2 (líneas divergentes). Como la mayoría de los espacios ROC están situados por encima de la diagonal y generalmente agrupados se puede concluir que la regresión simple por marcador es un método adecuado para estimar los parámetros de evaluación de la estructura genética. En términos generales, las simulaciones presentaron un rendimiento moderado en la discriminación entre falsos positivos y verdaderos positivos.

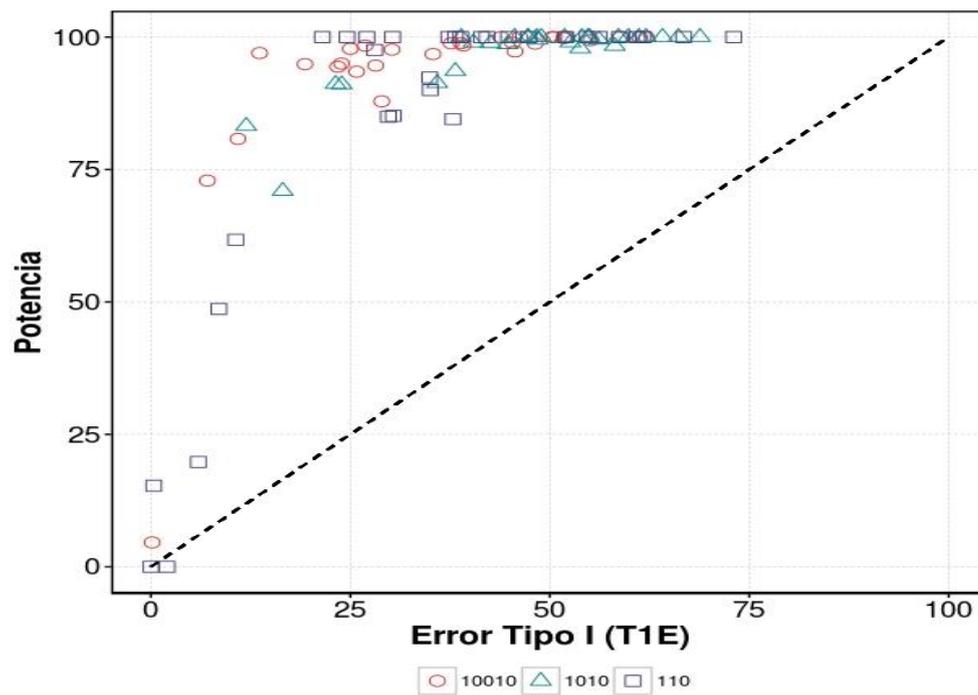


Figura 6. Gráfico ROC utilizando resultados predictivos de la matriz de confusión para calcular la potencia y el error Tipo I.

### Deriva Génica y Selección

El efecto de la deriva solo fue estudiado para la estrategia 2. La evaluación del efecto de la deriva en el estudio de simulación del experimento de selección divergente por capacidad uterina se abordó estudiando: (a) la varianza de la respuesta a la selección entre grupos, y (b) la respuesta a la selección y la mediana del número de QTN fijados en tres escenarios.

El efecto de la deriva génica se midió como la varianza entre las réplicas dentro de cada grupo. La varianza de la repuesta a la selección en los tres grupos de estructura genética (10010, 1010 y 110) fue de 0.52, 0.36 y 0.37, respectivamente. En el grupo de 10,010 QTN, la deriva génica fue 1.43 veces mayor que en el grupo con menor varianza (grupo B). Esto sugiere que en

poblaciones finitas, la deriva génica tendría mayor influencia en caracteres determinados por un mayor número de genes.

Para estudiar qué parte de la respuesta se debe a la selección y qué parte a la deriva se llevó a cabo una simulación donde se compararon tres escenarios de selección o apareamiento al azar durante 22 generaciones. En todos los escenarios se asumió un número de 10,010 QTN. Los escenarios son: (A) escenario 1: 22 generaciones de apareamientos al azar, (B) escenario 2: 12 generaciones de selección por tamaño de camada al destete en línea V más 10 generaciones de apareamientos al azar, y (C) escenario 3: 12 generaciones de selección por tamaño de camada al destete en línea V más 10 generaciones de selección divergente por capacidad uterina. Este escenario se corresponde con el experimento de selección divergente objeto de este estudio. Antes de comenzar a simular los tres escenarios, se simularon 10 generaciones de apareamiento al azar para simular la fundación de la línea V. Así, en total se simularon 32 generaciones. Se realizaron un total de 105 réplicas de simulación, 35 para cada escenario. En la Tabla 8 se presentan los resultados de los diferentes escenarios:

Tabla 8. Parámetros de estimación del efecto de la deriva génica en los tres escenarios para las líneas divergentes.

Línea	Escenario	Gen <sup>1</sup>	Varianza de Respuesta de Selección	Número de QTN-Fijados	
				Mediana	CV
Alta	(1) Apareamiento al Azar	22	0.0389	3,473	0.03
	(2) Selección NW <sup>2</sup> y Apareamiento al Azar	12 10	0.0178	7,217	0.06
	(3) Selección NW <sup>2</sup> y Selección UC <sup>3</sup>	12 10	0.2218	8,540	0.03
	(1) Apareamiento al Azar	22	0.0357	3,522	0.04
Baja	(2) Selección NW <sup>2</sup> y Apareamiento al Azar	12 10	0.0264	7,231	0.07
	(3) Selección NW <sup>2</sup> y Selección UC <sup>3</sup>	12 10	0.3298	8,405	0.04

Gen<sup>1</sup>: número de generaciones; NW<sup>2</sup>: tamaño de camada al destete; UC<sup>3</sup>: capacidad uterina.

La deriva se debe a que la elección de los hijos de los progenitores seleccionados es un proceso aleatorio. Por término medio, todos los hijos de un mismo padre y madre tienen el mismo valor mejorante, pero cada hijo tiene su propio valor debido a los eventos de recombinación y meiosis en el proceso de formación de gametos.

Los resultados en la Tabla 8 muestran que la varianza de respuesta en el escenario 2, es 2.18 y 1.35 veces mayor que la varianza del escenario 1 en las líneas alta y baja, respectivamente. Los efectos de la selección por tamaño de camada al destete hasta la generación 12 de la línea V condicionan que la varianza de la selección al final del proceso sea menor en el escenario 2 versus el 1, donde no hay selección. Esto es debido a que un gran número de QTN se habrán fijado en gran parte debido a la selección por el tamaño de camada (7,217 y 7,231 de mediana en las líneas alta y baja respectivamente). No obstante, los valores de 2.18 y 1.35 son pequeños si se comparan los escenarios 2 y 3. En efecto, el escenario 3 tiene 12.46 y 12.49 veces mayor varianza de respuesta a la selección que el escenario 2 para las líneas alta y baja, respectivamente. Además, en el escenario 3, la línea baja presentó 1.49 veces más varianza de respuesta que en la línea alta. La menor respuesta encontrada en la línea alta podría ser debida a la mayor número de QTNs con efecto favorable que presentan frecuencias elevadas debido a la previa selección por tamaño de camada, lo que conllevaría una menor varianza.

Con respecto al número de QTNs fijados, para la selección en la línea alta, un 40% de estos fueron resultado de los efectos de la deriva (escenario 1) y un 15.49% fueron fijados por efecto de la selección de capacidad uterina más la deriva. Esto indica que la deriva y la selección por tamaño de camada al destete contribuye con 84.50% de los QTNs fijados del escenario 3. Además, puede observarse en la Tabla 8 que en la línea baja un 42% de QTN se encontraron fijados por deriva, mientras que el 13.97% fuera resultado del proceso de selección de capacidad uterina más la deriva, y el otro 86.03% corresponde a QNTs fijados por la selección en el escenario 2. Tomando los valores mínimos y máximos encontrados en la historia de la población, se encuentra que entre un 35 al 50% de QTN fijados se deben al efecto de la deriva, siendo más pronunciado en la línea baja (hasta 50%).

En todo caso, los resultados indican que al existir una gran cantidad de QTN con efectos de magnitudes parecidas (escenario 1), pueden fijarse alelos contrarios en diferentes QTNs, lo cual contribuye con un efecto pequeño sobre la respuesta a la selección (modelo infinitesimal). En efecto, la Figura 7 muestra los diferentes escenarios simulados. De modo que la deriva representa por término medio el 14.5% (comparando con el escenario 1) y 10.5% (comparando con el escenario 2) de la respuesta obtenida por la selección por capacidad uterina en la línea alta. Para la línea baja los valores de la deriva en la selección por capacidad uterina fueron de 12.7% y 10.2% comparados con los escenarios 1 y 2, respectivamente. Estos valores fueron calculados teniendo repuestas promedios de 1.676 y 1.910 en las líneas alta y baja, respectivamente.

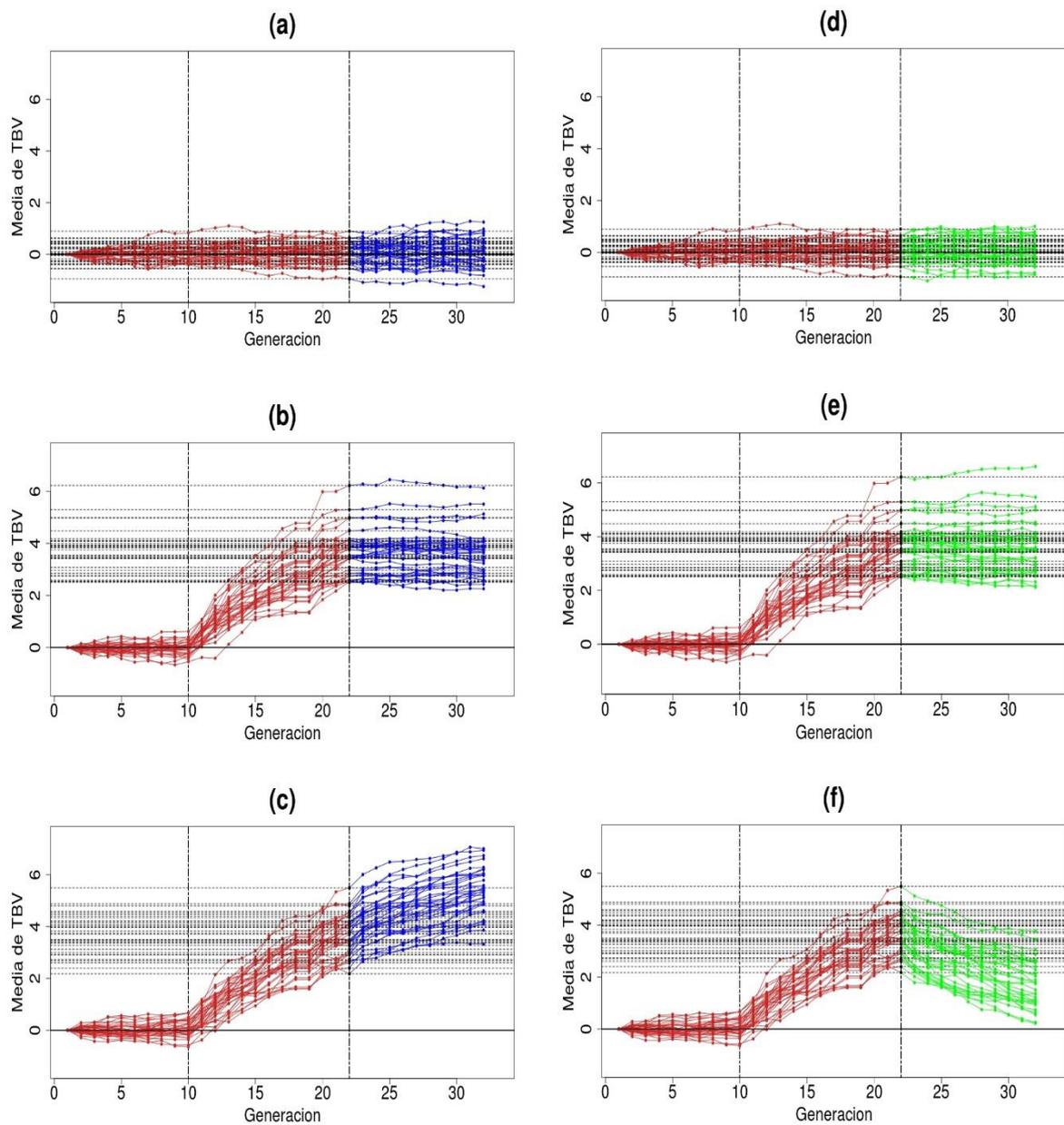


Figura 7. Gráficos de las medias del verdadero valor mejorante (TBV) por generación en diferentes escenarios de deriva génica y selección. Las medias del verdadero valor mejorante (TBV) son la respuesta a la selección por generación. Los escenarios son: (a y d) apareamientos al azar; (b y e) selección por tamaño de camada al destete más apareamiento al azar; (c y f) selección por tamaño de camada al destete más selección por capacidad uterina. Las líneas rojas representan los valores de TBV de capacidad uterina antes de iniciar el experimento de selección divergente (generación del 0 al 22). Las líneas horizontales discontinuas de color negro representan los diferentes puntos de partida de la selección por capacidad uterina. Los tres gráficos a la izquierda pertenecen a la población de la línea alta (a, b y c), mientras que los tres gráficos a la derecha pertenecen a la línea baja (d, e y f). La línea de color azul es la respuesta a la selección por capacidad uterina en la línea alta (generación del 23 a 32) y la línea de color verde es la respuesta a la selección por capacidad uterina en la línea baja (generación del 23 a 32).

## Discusión

En la discusión se usarán los términos QTL o QTN como términos similares en función de su definición como sitio o área genómica que explica la varianza genética de un carácter cuantitativo. En las simulaciones creadas en nuestra investigación no hubo una asignación fija e invariable de la ubicación de los QTNs. Las posiciones no fueron fijadas para evitar resultados dependientes de este factor, porque se desconoce la ubicación de los QTNs de la capacidad uterina en el genoma del conejo. No obstante, las simulaciones funcionaron correctamente para producir el experimento de selección divergente por capacidad uterina.

### Análisis por la Fijación de QTNs

En la comparación entre las estrategias de las líneas divergentes y la línea única para QTNs fijados y no fijados, se demuestra que tener menor número de QTN fijados y también un número mayor de QTNs con valores altos de frecuencia intermedia produce una diferencia a favor de la estrategia de utilizar líneas divergentes frente a utilizar la línea única. El aumento de los QTNs de frecuencias intermedias en un área genómica, aumentaría la asociación con los SNPs cercanos a los QTNs dentro de cada línea, lo cual provocaría una diferenciación genética en esta área y por lo tanto aumentaría la probabilidad de detección del QTN. Esta situación se presenta en la estrategia de líneas divergentes y no en la línea única. Esto parece ocurrir en los datos reales de estudios de diferenciación genética en pollo, utilizando un rango de 10 a 49 animales en las generaciones 40 y 50 de líneas divergentes para el carácter peso vivo al sacrificio, se logró determinar huellas de la selección sobre el genoma, particularmente en la región 60-80 Mb del cromosoma 4. De acuerdo a estos trabajos, después de 50 generaciones de selección se siguen encontrando nuevos sitios de diferenciación genética debido al efecto de la selección divergente, mediante el análisis de las frecuencias alélicas entre las líneas divergentes. (Johansson et al., 2010). En gallinas ponedoras, en poblaciones divergentes de 11 generaciones de selección, también fue posible distinguir regiones genómicas con pocos individuos (41 en línea alta y 34 en la línea baja), para un carácter con heredabilidad baja como es el registro del comportamiento agresivo de picotazos (Grams et al., 2015). Por lo tanto, el muestreo de 160 individuos de líneas divergentes podría ser suficiente para encontrar asociaciones positivas de interés que permitan conocer mejor la estructura genética del carácter capacidad uterina y por tanto del tamaño de camada.

## Comparación de Grupos de Estructura Genética

### *Criterios de Determinación de la Asociación*

El objetivo de esta investigación fue dilucidar si un carácter complejo como el tamaño de camada mostraría alguna asociación en un GWAS con un número de individuos limitado. Los resultados demuestran que es posible encontrar asociaciones positivas pero con un elevado número de falsos positivos. El análisis del estudio utilizó una metodología clásica del GWAS, basada en una regresión lineal simple por marcador (Sham y Purcell, 2014).

Los límites de los criterios utilizados para decidir si una asociación es verdadera o falsa siguen siendo un problema, porque estos criterios incluyen un fuerte componente de arbitrariedad. Los criterios utilizados en nuestra investigación fueron la distancia genómica y el desequilibrio de ligamiento (LD) utilizando  $r^2$ . En humanos en donde el desequilibrio de ligamiento es bajo, las variante causales y los SNPs asociados tienen valores de  $r^2$  entre 0.017 a 0.378, siendo más común valores inferiores de 0.10 pero utilizando grandes tamaños de muestra (Wang et al., 2010; Zhu et al., 2012). En cambio, en animales el patrón de desequilibrio de ligamiento es alto, por ejemplo en el genoma porcino se han encontrado en 1 Mb valores promedio de  $r^2$  entre 0.20 a 0.30 (Du, et al. 2007; Wang et al. 2013). En conejos en la raza Neozelandesa, el desequilibrio de ligamiento permanece mayor que 0.30 entre SNP a una distancia de 1.20 Mb, inclusive este valor puede ser constante en bloques de desequilibrio de ligamiento de 3.2 Mb (Carneiro et al., 2011). El utilizar un valor de  $r^2$  de 0.5 como medida de desequilibrio de ligamiento fue muy conservadora para determinar una asociación verdadera. La potencia de detección estimada previo al experimento fue muy baja, entre un 10 a 30% para 160 animales con un  $r^2$  de 0.5, lo cual concuerda con la no detección de asociaciones en una línea única (Hayes, 2013). No obstante, si con el valor de 0.5 se consigue una buena potencia de detección con los datos simulados usando líneas divergentes, entonces en los datos reales sería previsible tener más éxito.

En cuanto a la distancia, la decisión de 1 Mb en vez de una cantidad mayor se sostiene por la reducción del 30 al 40% en la precisión cuando se utilizan distancias iguales o mayores a 5 Mb para QTL que explican del 10% - 18% de la varianza fenotípica (MacLeod et al., 2010). MacLeod et al. (2010) encontraron en sus simulaciones que el SNP más próximo no es el más asociado al QTL. No obstante, en nuestro estudio al tener una buena densidad de SNPs, fue posible testar la asociación de un QTN con varios SNPs.

### ***Error Tipo I (T1E)***

En nuestra investigación, los promedios del error tipo I fueron de 37%, 48% y 36% para los grupos de QTNs encontrados en nuestro estudio, que difieren con los encontrados en otro estudio de simulación por regresión simple por marcador, que apenas fueron del 1% (De Maturana et al., 2014). Empero, ellos utilizaron como SNP el propio QTN, una distribución gamma para los efectos de QTNs (grandes y pequeños), y valores predefinidos de desequilibrio de ligamiento y menor frecuencia alélica. Además, nosotros utilizamos en promedio 15,000 marcadores para estimar asociaciones con 160 animales; en contraste, la investigación De Maturana et al. (2014) uso un número de muestra alto (1000 individuos) comparado con sus 320 marcadores, por lo tanto su estimación fue más exacta. También, otra de las razones que pudo explicar la moderada tasa de error tipo I de nuestro resultado fue la estratificación de la población al utilizar individuos de ambas líneas (Balding, 2006; Hayes, 2013) o la exclusión del pedigrí, lo cual produce 1.38 veces más falsos positivos (MacLeod et al., 2010).

### ***La Potencia de Detección***

En cuanto a la potencia, nuestros resultados con una línea única son acordes con las simulaciones de Morin et al. (2009) en donde demuestra que es muy difícil obtener suficiente potencia para detectar asociaciones entre marcadores y QTNs. Sin embargo, nuestro trabajo presentó una potencia muy alta utilizando líneas divergentes, inclusive en la estructura con 110 QTN (83%). Nuestros resultados fueron superiores a los De Maturana et al. (2014), en la cual encontraron valores entre el 65 al 90% para QTNs de efecto grande y solo de 0.8 a 5.6% para QTNs de efecto pequeño. MacLeod et al. (2010) encontró que la potencia de detección de un QTL se incrementa con una menor frecuencia alélica (MAF) mayor de 0.30, llegando a obtener hasta 70% de potencia. Esto concuerda con los valores altos para el grupo de 1,010 y 10,010 QTNs, en la cual más del 80% de los QTNs tenían un MAF >0.35. En otras palabras, los valores de potencia de nuestro estudio son altos en la mayoría de los casos porque los QTN tenían un MAF relativamente alto y solo se incluyó QTNs no fijados.

### ***Valor Predictivo Predicho***

El promedio de los grupos de 1,010 y 10,010 QTN del valor predictivo positivo (PPV) fue de 29% y 57%, datos dentro del rango de resultados obtenidos por De Maturana et al. 2014 (0.29 a 0.57). Un alto desequilibrio de ligamiento entre QTNs de grande y pequeño efecto produce que solo sea detectado el QTN con mayor efecto, especialmente si ellos tienen valores de MAF bajos. La mayoría de los QTN de nuestra investigación fueron detectados por su elevado MAF usando las líneas divergentes. Esto concuerda con el experimento de De Maturana et al.

(2014), que mostro que QTNs con alto MAF en cualquier escenario pueden ser detectados, por el contrario una MAF bajo afectaría la detección. La mayoría de los QTNs de nuestros estudios fueron detectados a través de la asociación de los SNPs por regresión simple por marcador, pero no todas las asociaciones positivas estimadas cumplían con los criterios asignados por nosotros para determinarlas como una asociación verdadera. Por tal motivo, el grupo de 110 QTN obtuvo un bajo valor predictivo positivo (8%).

En términos generales, valores por encima del 80% de potencia son aceptables en los estudios de GWAS (Klein, 2007) y se espera que el error tipo I no supere la probabilidad fijada al principio del experimento ( $P < 0.05$ ). En nuestro caso con 160 animales de las líneas divergentes, obtuvimos una gran potencia, pero también un elevado error tipo I.

### **Deriva Génica y Selección**

La deriva génica generalmente suele medirse en genética de poblaciones por los cambios de frecuencias alélicas y los cambios en la varianza genética. En nuestra investigación, la estructura genética con mayor número de QTNs obtuvo el mayor impacto de la deriva génica. En nuestro caso se puede observar que después de las 12 generaciones de selección en la línea V, una gran parte de QTNs se encuentran fijados por la acción de ambas fuerzas evolutivas (7217 en la línea alta y 7231 en la línea baja), pero nosotros asumimos una distribución uniforme de las frecuencias iniciales. La deriva provocó la fijación de QTNs dentro de un rango del 35 al 50% de QTNs de la selección por capacidad uterina. En la investigación de Johansson et al. (2010) con una asignación inicial fija de SNPs a frecuencias extremas, la deriva provocó hasta un 50% de fijación de SNP, similar a la de nuestro estudio; sin embargo, cuando ellos asignaron frecuencias intermedias iniciales, estas generaron un 20% de fijación.

Los resultados de evaluar la deriva en las líneas divergente sugieren que esta tiene menor efecto en la línea baja (12.7%) por término medio en la respuesta a la selección. Esto se debe que los QTNs no fijados estaban a frecuencias extremas después de la selección por tamaño de camada al destete en la línea V. Por consiguiente, el efecto de la selección en la línea alta, previsiblemente fuera menor que en la línea baja y por ende un efecto mayor de la deriva génica. En la línea alta la deriva fue 14.5% de la respuesta de selección por capacidad uterina.

La deriva en nuestras simulaciones presentó un efecto pequeño sobre la respuesta a la selección. Además, nuestros resultados del efecto de deriva fueron similares al experimento de Grams et al. (2015), lo cuales demostraron a través de simulaciones muy conservadoras con

frecuencias intermedias iniciales, que la respuesta al carácter era el resultado conjunto de deriva y selección, con una mayor proporción de la selección.

Esta investigación basó la complejidad del tamaño de camada en los efectos posibles de la deriva y la selección divergente. Sin embargo, efectos de dominancia y epistasis podrían influir en el tamaño de camada (Nagy et al. 2014 en conejo; Peripato et al. 2004 en ratón). Este experimento se desarrolló de esta forma por las limitaciones en el conocimiento sobre la estructura genética específica para el tamaño de camada y a expensas de la inexistencia de estudios de GWAS para caracteres reproductivos en conejo. Por lo tanto, para futuras investigaciones se considerará la interacción gen-gen y diferentes métodos de asociaciones de múltiples marcadores con la información obtenida del análisis genómico de datos reales.

## **Conclusiones**

La estrategia de utilizar una cantidad moderada de individuos procedentes de dos poblaciones seleccionadas de forma divergente podría aumentar la potencia de detección en un GWAS frente a la alternativa de utilizar la misma cantidad de individuos de una población seleccionada. Si el modelo infinitesimal fuera el apropiado para la estructura genética real del carácter capacidad uterina, los resultados serían similares al grupo de 10,010 QTNs. Este grupo presenta un mayor número de valor predictivo positivo, pero un elevado error tipo I. Por consiguiente, con datos reales de líneas divergentes podrían aparecer abundantes asociaciones a lo largo del genoma pero produciendo poca discriminación entre los verdaderos y los falsos positivos. Además, se infiere que la deriva génica presentó poco efecto sobre la respuesta a la selección.

## **Agradecimientos**

El trabajo ha sido financiado por la Secretaría de Estado I+D+i del Ministerio de Economía y Competitividad de España (AGL2014-55921-C2-1-P), y B. S. Sosa-Madrid por la Secretaría Nacional de Ciencia, Innovación y Tecnología (SENACYT) de la República de Panamá.

## **Referencias Bibliográficas**

- Balding, D. (2006). A tutorial on statistical methods for population association studies. *Nature Reviews Genetics* 7, 781–791.
- Blasco, A., Ortega, J.A., Climent, A., y Santacreu, M.A. (2005). Divergent Selection for Uterine Capacity in Rabbits . I . Genetic Parameters and Response to Selection. *Journal*

- Animal Science 83, 2297–2302.
- Carneiro, M., Afonso, S., Geraldes, A., Garreau, H., Bolet, G., Boucher, S., Tircazes, A., Queney, G., Nachman, M.W., y Ferrand, N. (2011). The Genetic Structure of Domestic Rabbits. *Molecular Biology and Evolution* 28, 1801–1816.
- Chen, G.K., Marjoram, P., y Wall, J.D. (2009). Fast and flexible Simulation of DNA Sequence Data. *Genome research* 19, 136–142.
- Demars, J., Fabre, S., Sarry, J., Rossetti, R., Gilbert, H., Persani, L., Tosser-Klopp, G., Mulsant, P., Nowak, Z., Drobik, W., et al. (2013). Genome-Wide Association Studies Identify Two Novel BMP15 Mutations Responsible for an Atypical Hyperproliferacy Phenotype in Sheep. *PLoS Genetics* 9, e1003482.
- Du, F.-X., Clutter, A.C., y Lohuis, M.M. (2007). Characterizing linkage disequilibrium in pig populations. *International journal of biological sciences* 3, 166–178.
- Estany, J., Baselga, M., Blasco, A., y Camacho, J. (1989). Mixed Model methodology for Estimation of genetic Response to Selection in Litter Size of Rabbits. *Livestock Production Science* 21, 67–75.
- Falconer, D.S., y Mackay, T.F.C. (1996). *Introducción a la Genética Cuantitativa* (Zaragoza, España: Editorial ACRIBIA).
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters* 27, 861–874.
- García, M.L., y Baselga, M. (2002). Estimation of Genetic Response to selection in litter size of rabbits using a cryopreserved control population. *Livestock Production Science* 74, 45–53.
- Grams, V., Wellmann, R., Preuß, S., Grashorn, M.A., Kjaer, J.B., Bessei, W., y Bennewitz, J. (2015). Genetic parameters and signatures of selection in two divergent laying hen lines selected for feather pecking behaviour. *Genetics Selection Evolution* 47, 77.
- Hayes, B. (2013). Overview of statistical methods for genome-wide association studies (GWAS). In *Genome-Wide Association Studies and Genomic Prediction*, C. Gondro, J.H. van der Werf, y B. Hayes, eds. (Humana Press), pp. 149–169.
- Hickey, J.M., y Gorjanc, G. (2013). Simulated Data for Genomic Selection and Genome-Wide Association Studies Using a Combination of Coalescent and Gene Drop Methods. *G3 Genes | Genomes | Genetics* 2, 425–427.
- Johansson, A.M., Pettersson, M.E., Siegel, P.B., y Carlborg, Ö. (2010). Genome-Wide Effects of Long-Term Divergent Selection. *PLoS Genetics* 6, e1001188.
- Klein, R.J. (2007). Power analysis for genome-wide association studies. *BMC genetics* 8, 58.
- MacLeod, I.M., Hayes, B.J., Savin, K.W., Chamberlain, A.J., McPartlan, H.C., y Goddard, M.E.

- (2010). Power of a genome scan to detect and locate quantitative trait loci in cattle using dense single nucleotide polymorphisms. *Journal of Animal Breeding and Genetics* 127, 133–142.
- De Maturana, E.L., Ibáñez-Escriche, N., González-Recio, Ó., Marenne, G., Mehrban, H., Chanock, S.J., Goddard, M.E., y Malats, N. (2014). Next generation modeling in GWAS: comparing different genetic architectures. *Human Genetics* 133, 1235–1253.
- Morin, P.A., Martien, K.K., y Taylor, B.L. (2009). Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources* 9, 66–73.
- Nagy, I., Farkas, J., Curik, I., Gorjanc, G., Gyovai, P., y Szendrő, Z. (2014). Estimation of additive and dominance variance for litter size components in rabbits. *Czech Journal of Animal Science* 59, 182–189.
- Onteru, S.K., Fan, B., Du, Z.-Q., Garrick, D.J., Stalder, K.J., y Rothschild, M.F. (2011). A whole-genome association study for pig reproductive traits. *Animal Genetics* 43, 18–26.
- Peripato, A.C., De Brito, R.A., Matioli, S.R., Pletscher, L.S., Vaughn, T.T., y Cheverud, J.M. (2004). Epistasis affecting litter size in mice. *Journal of Evolutionary Biology* 17, 593–602.
- Purcell, S., Cherny, S.S., y Sham, P.C. (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics (Oxford, England)* 19, 149–150.
- Schneider, J.F., Nonneman, D.J., Wiedmann, R.T., Vallet, J.L., y Rohrer, G.A. (2014). Genomewide association and identification of candidate genes for ovulation rate in swine. *Journal of Animal Science* 92, 3792–3803.
- Sham, P.C., y Purcell, S.M. (2014). Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics* 15, 335–346.
- Stram, D.O. (2014). *Design, Analysis, and Interpretation of Genome-Wide Association Scans* (New York, NY: Springer New York).
- Wang, K., Dickson, S.P., Stolle, C.A., Krantz, I.D., Goldstein, D.B., y Hakonarson, H. (2010). Interpretation of Association Signals and Identification of Causal Variants from Genome-wide Association Studies. *American Journal of Human Genetics* 86, 730–742.
- Wang, L., Sørensen, P., Janss, L., Ostersen, T., y Edwards, D. (2013). Genome-wide and local pattern of linkage disequilibrium and persistence of phase for 3 Danish pig breeds.

BMC genetics *14*, 115.

Zhu, Q., Ge, D., Heinzen, E.L., Dickson, S.P., Urban, T.J., Zhu, M., Maia, J.M., He, M., Zhao, Q., Shianna, K. V., et al. (2012). Prioritizing genetic variants for causality on the basis of preferential linkage disequilibrium. *American Journal of Human Genetics* *91*, 422–434.

## **ARTÍCULO 2: ESTUDIO DE ASOCIACIÓN DEL GENOMA COMPLETO Y HUELLAS DE SELECCIÓN EN LINEAS DIVERGENTES SELECCIONADAS PARA CAPACIDAD UTERINA EN CONEJO.**

### **Resumen**

Los principales factores que repercuten en la economía de la producción de carne de conejo son la alimentación y la reproducción. La mejora del tamaño de camada y sus componentes contribuye a la eficiencia productiva basada en el aumento del número de gazapos por hembra y parto. Los dos componentes del tamaño de camada más importantes son la tasa de ovulación y la supervivencia prenatal. En condiciones de atestamiento uterino, la supervivencia prenatal que depende de la hembra es la capacidad uterina. Con el objetivo de entender la base genética del tamaño de camada, se realizó un análisis de asociación de genoma completo y la estimación del índice de diferenciación genética ( $F_{st}$ ) en tres líneas, dos líneas seleccionadas de forma divergente por capacidad uterina (UC) durante diez generaciones de selección y una población de referencia (línea V). Muestras de sangre de 183 hembras recolectadas de las generaciones de selección relajada 11 y 12 fueron utilizadas para obtener datos genómicos de 200K Affymetrix Axiom OrcunSNP Array. Los caracteres registrados en estas hembras fueron la tasa de ovulación (OR), el número de embriones implantados (IE), la supervivencia embrionaria (ES), la supervivencia fetal (FS), la supervivencia prenatal (PS), el tamaño de camada o nacidos totales (LS), el número de nacidos vivos (NBA) y nacidos muertos (NBD) de la segunda gestación. Un total de 114,659 SNPs fueron utilizados para la estimación de cada efecto de SNP a través de un Bayes B con GenSel Software. Los SNPs fueron analizados conjuntamente a través de 2,171 ventanas de 1-Mb y se estimó la proporción de la varianza genética. Aquellas ventanas que superaron el 1% de la varianza genética fueron consideradas como QTLs putativos. Además, aquellas áreas genómicas formando un “clúster” de SNPs relevantes, con un valor de  $F_{st} > 0.4$ , se consideraron como huellas de la selección. Para los caracteres LS, OR, IE y NBA, se encontraron 4, 1, 5 y 4 regiones asociadas, respectivamente. Los demás caracteres (ES, FS, PS y NBD) no presentaron QTLs putativos. Se encontraron dos ventanas en el cromosoma 17 asociadas a los caracteres reproductivos LS, IE y NBA. Estos dos QTLs explicaban hasta un 19.47% de la varianza genómica del tamaño de camada. El gen ERO1A reportado en un estudio previo de análisis de expresión en estas líneas de conejo, se encontró dentro de uno de estos dos QTLs. Se detectaron huellas de selección por  $F_{st}$  entre la línea baja y la línea de referencia pero sus localizaciones son diferentes del GWAS. Además, se detectaron con el GWAS áreas genómicas no reportadas anteriormente. En estas

áreas están localizados genes como por ejemplo BMP4, PTGDR ANKH y CDKN3 cuyas funciones biológicas tienen que ver con la actividad de receptores de prostanoides, con componentes asociadas al citoplasma y con la regulación negativa de la proliferación celular. Estas áreas deberán ser estudiadas con más detalle y ser validadas en las poblaciones seleccionadas por caracteres maternales antes de ser utilizadas para mejorar la respuesta a la selección de los caracteres reproductivos en los programas de mejora.

**Palabras Claves:** GWAS, capacidad uterina, conejo, tamaño de camada, Fst, huellas de selección, divergencia.

## Introducción

En las especies prolíficas, los caracteres reproductivos tienen una gran importancia económica, en especial el tamaño de camada, porque los costes fijos se disminuyen con el incremento del número de descendientes por hembra (Baselga y Blasco, 1989). Sin embargo, la respuesta para el tamaño de camada en los experimentos de selección ha sido baja, alrededor de 0.1 gazapos por generación (García y Baselga, 2002a; Khalil y Al-Saef, 2008). Esta situación ha llevado a la búsqueda de alternativas para intentar mejorar el tamaño de camada de forma más eficaz. Una de las alternativas propuestas fue seleccionar por los caracteres que determinan el tamaño de camada (Argente et al., 1997; Laborda et al., 2012; Ziadi et al., 2013).

Los principales componentes del tamaño de camada son la tasa de ovulación y la supervivencia prenatal (Khalil y Al-Saef, 2008; Ragab et al., 2014). Cuando la tasa de ovulación no es un factor limitante y se atesta el útero de embriones, puede medirse la capacidad uterina como el número de nacidos totales al parto (Christenson et al. 1987; Mocé et al. 2004). La técnica que se ha utilizado en conejo para atestar el útero es la ovariectomía unilateral. Esta técnica consiste en la extirpación de uno de los ovarios. Al extirpar un ovario se produce un fenómeno de compensación ovárica en el único ovario presente, que ovula igual que la media de los dos ovarios de una hembra intacta (Argente et al. 1997; Blasco et al. 1994). Como los dos cuernos uterinos desembocan de forma independiente en el cérvix, no hay transmigración de embriones y el cuerno uterino adyacente al ovario queda atestado y se puede expresar la capacidad uterina.

Entre 1991 y 1998 se llevó a cabo un experimento de selección divergente por capacidad uterina en el Departamento de Ciencia Animal de la Universidad Politécnica de Valencia. La

capacidad uterina se estimó como el tamaño de camada en hembras unilateralmente ovariectomizadas (hembras ULO). Al principio del experimento se vitrificaron embriones para luego ser utilizados para reconstituir una población de referencia. Después de diez generaciones de selección, la divergencia entre las dos líneas fue de 1.5 gazapos para la capacidad uterina (Blasco et al., 2005), con una respuesta correlacionada para el tamaño de camada de 2.35 gazapos (Santacreu et al., 2005). La correlación genética entre el tamaño de camada y la capacidad uterina fue elevada, 0.92 (Argente et al., 2000). El estudio sobre qué regiones genómicas y qué posibles genes podrían explicar la diferencia en tamaño de camada de las líneas divergentes es de gran interés para conocer la naturaleza genética del tamaño de camada. Actualmente en conejo se dispone de un panel de densidad de SNPs de 200K Affymetrix Axiom OrcunSNP Array y una base de datos del ensamblaje de las anotaciones genómicas, *OryCun2.0* que hace posible el estudio genómico de las líneas del experimento de selección divergente por capacidad uterina.

El objetivo de este estudio fue identificar regiones genómicas y genes candidatos putativos asociados al carácter tamaño de camada y otros caracteres reproductivos para incrementar la información disponible sobre posibles QTLs involucrados en su determinación biológica. Además, otro objetivo fue determinar las huellas de selección mediante la estimación de los índices de diferenciación genética en las tres líneas utilizadas en el estudio.

## **Material y Métodos.**

Este estudio fue aprobado por el Ministerio de Economía y Competitividad de España y por el Comité Ético de la Universidad Politécnica de Valencia.

### **Animales**

Los animales de este trabajo proceden de un experimento de selección divergente por capacidad uterina desarrollado desde el año 1991 a 1998. La capacidad uterina fue evaluada durante la selección como el tamaño de camada en hembras unilateralmente ovariectomizadas. La ovariectomía unilateral se realizó antes de alcanzar la pubertad. Las conejas estuvieron en un sistema de producción donde se llevaban a su primera cubrición a los 4.5 meses edad, se destetaban los gazapos a los 28 días de lactación y el intervalo parto-cubrición era alrededor de 10 días (Blasco et al., 2005). La selección se realizó bajo un modelo animal de medidas repetidas utilizando BLUP para todos los animales, los machos se seleccionaron dentro de familias de machos para el control de la consanguinidad (Santacreu et al., 2005). El experimento de selección divergente por capacidad uterina está descrito con mayor detalle por

Argente et al. (1997). Después de diez generaciones de selección divergente por capacidad uterina, la selección fue relajada durante seis generaciones para estimar la respuesta correlacionada del tamaño de camada. Para el experimento se recogió sangre de 90 hembras intactas de la línea seleccionada para aumentar la capacidad uterina (línea Alta), 69 hembras intactas de la línea seleccionada para disminuir la capacidad uterina (línea Baja) y 30 hembras de una población de referencia (línea V). Las hembras de las líneas alta y baja provenían de las generaciones relajadas 11 y 12 (Blasco et al., 2005; Mocé et al., 2005). La población base de las líneas divergentes de capacidad uterina proviene de la generación 12 de una línea seleccionada por tamaño de camada al destete (la línea V). Para obtener una línea de referencia se conservaron embriones vitrificados de las generaciones 13 y 15 de la línea V. Después de 10 generaciones de selección por capacidad uterina, estos embriones fueron descongelados y transferidos a hembras receptoras para reconstituir esas generaciones. De estos embriones se obtuvieron hembras y machos, los cuales se aparearon al azar dentro de generación para obtener la población de referencia (Santacreu et al., 2005).

### **Caracteres**

Para este trabajo se utilizaron los datos de la segunda gestación de las conejas intactas de las líneas alta, baja y línea V. A los doce días de la segunda gestación se contó por laparoscopia la tasa de ovulación (OR, estimada como el número de cuerpos lúteos presentes en ambos ovarios) y el número de embriones implantados en ambos cuernos uterinos (IE). Al parto se registró el tamaño de camada al nacimiento o número de nacidos totales (LS), nacidos vivos (NBA) y nacidos muertos (NBD). Además se calculó la supervivencia embrionaria (ES) como el cociente  $IE/OR$ , la supervivencia fetal (FS) como  $LS/IE$  y la supervivencia prenatal (PS) como  $LS/OR$ . El resumen estadístico de los datos del experimento se muestra en la Tabla 1.

Tabla 1. Análisis descriptivo del tamaño de camada y sus componentes.

Carácter	Abre. <sup>1</sup>	N <sup>2</sup>	Media	SD <sup>3</sup>	Min <sup>4</sup>	Max <sup>5</sup>
Tamaño de Camada	LS	183	8.88	3.16	1.00	17.00
Número de Nacidos Vivos	NBA	183	8.26	3.35	0.00	15.00
Número de Nacidos Muertos	NBD	183	0.62	1.68	0.00	11.00
Tasa de Ovulación	OR	157	14.84	2.52	9.00	22.00
Embriones Implantados	IE	158	12.15	2.98	3.00	19.00
Supervivencia Embrionaria	ES	154	0.82	0.17	0.25	1.00
Supervivencia Fetal	FS	158	0.75	0.19	0.09	1.00
Supervivencia Prenatal	PS	157	0.62	0.21	0.06	1.00

<sup>1</sup> Abre.: Abreviación; <sup>2</sup> N: Número de Registros; <sup>3</sup> SD: Desviación Estándar; <sup>4</sup> Min: Mínimo; <sup>5</sup> Max: Máximo

### Extracción de DNA, Genotipado y Control de Calidad.

El DNA genómico fue extraído de muestras congeladas de sangre usando el Kit de Favorgen (FABGK 001-2). Se utilizaron 189 muestras con concentraciones mínimas de 20 ng/ $\mu$ l y un volumen mínimo de 45  $\mu$ l. La concentración fue estimada con Nanodrop ND-1000 y confirmada por PicoGreen antes de empezar el genotipado. Para la calidad e integridad de DNA se utilizaron muestras con un ratio de OD<sub>260</sub>/OD<sub>280</sub> entre 1.8-2.0, un ratio OD<sub>260</sub>/OD<sub>230</sub> superior a 1.5 y una comprobación de la integridad en gel de agarosa 1%. La extracción de DNA se realizó en el Departamento de Ciencia Animal de la Universidad de Lleida y el genotipado se realizó en el CEGEN (Centro Nacional de Genotipado) de la Universidad de Santiago de Compostela. Después de la extracción, dos muestras fueron excluidas por su baja integridad.

Los resultados fueron analizados con el software Axiom Analysis Suite 1.1.1.66 de Affymetrix. Los SNPs presentes fueron 199,692. Se utilizaron un conjunto de algoritmos de genotipado para analizar una gran cantidad de datos brutos y estimar la probabilidad que el genotipo fuera un heterocigoto (0), homocigoto "A" (1) o homocigoto "a" (2). Cada SNP debía sobrepasar los umbrales pre-definidos del software para ser considerados como un genotipo aceptable (Anderson et al., 2010).

El control de calidad de las muestras y los SNPs después del genotipado fue realizado utilizando dos programas Axiom Analysis Suite y ZANARDI. Los parámetros de calidad para las muestras fueron una tasa de genotipado (“call rate” - CR)  $\geq 97\%$ , DQC (Control de Calidad del Panel – “Dish Quality Control”)  $\geq 0.89$  y Plate QC (Control de Calidad de los Paneles)  $\geq 0.96$  con un promedio de muestras con CR de 99%. Además se realizó un test de inconsistencia mendeliana o parentesco y se eliminó una muestra que se designó como una réplica ya que poseía un valor menor de 1.5 y un 99.87% de similitud con otra muestra.

Para los SNPs se utilizó un tasa de genotipado (CR)  $> 0.95$ , genotipos faltantes  $< 0.97$  y una menor frecuencia alélica (MAF)  $> 0.03$ . Además aquellos SNPs localizados en el cromosoma sexual “Y”, y aquellos SNPs no asignados a ningún cromosoma en el genoma de referencia del conejo *OryCun2.0* ensamblado, (<http://www.ncbi.nlm.nih.gov/projects/mapview/> y [http://www.ensembl.org/Oryctolagus\\_cuniculus/Info/Index](http://www.ensembl.org/Oryctolagus_cuniculus/Info/Index)) también fueron eliminados. En la Tabla 2 se muestran los criterios de exclusión del control de calidad entre los que se incluye un test de equilibrio Hardy Weinberg de p-valor  $< 10^{-6}$ , comúnmente utilizado, y la categoría del SNP. Aquellos SNPs clasificados como PolyHighResolution y NoMinorHom no fueron excluidos.

El genotipado fue obtenido en formato Affymetrix pero convertido a formato Plink ACTG por el programa ZANARDI (Marras et al., 2015). Algunos genotipados faltantes en las muestras fueron imputados usando el procedimiento de BEAGLE v4.0 y solo los SNPs con una puntuación de calidad para la imputación de  $R^2 > 0.5$  fueron mantenidos como un genotipado válido (Browning y Browning 2009).

Tabla 2. Número de SNPs excluidos por diferentes criterios

Criterio	Número de SNP	Porcentaje
MonoHighResolution	17,282	8.65%
Otros	16,606	8.32%
CallRateBelowThreshold	7,910	3.96%
OTV (Off-Target Variant)	1,599	0.80%
No Asignados en Mapa Genómico	31,577	15.81%
Hardy Weinberg ( $p < 10^{-6}$ )	10,920	5.47%
Menor Frecuencia Alélica (MAF $< 0.03$ )	32,685	16.37%

Después del filtrado, un total de 114,659 SNPs (57%) permanecieron en el genotipo. En cuanto a las muestras, de las 189 muestras originales se utilizaron para el GWAS 183 datos.

### Estudio de Asociación del Genoma Completo

A fin de considerar una posible estratificación de la población, un gráfico de escalado multidimensional (MDS) de matriz por identidad de estado generó los dos componentes principales más importantes de la variación genética. Con este gráfico se intentó visualizar la existencia de distancias genéticas debido a la divergencia por el efecto de la selección y tener en cuenta el efecto de tres líneas en el modelo (Bergfelder-Drüing et al., 2015).

La asociación entre los SNPs y fenotipos para todos los caracteres fue obtenida utilizando el método de Bayes B, el cual analiza todos los SNPs simultáneamente (Método de Múltiples Marcadores) y asume diferentes varianzas genéticas para cada locus. El modelo estadístico utilizado fue:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \sum_{j=1}^k \mathbf{z}_j \alpha_j \delta_j + \mathbf{e}$$

En donde  $\mathbf{y}$  es el vector de valores fenotípicos,  $\mathbf{X}$  es la matriz de incidencia de efectos ambientales predefinidos,  $\mathbf{b}$  es un vector de efectos ambientales predefinidos (año-estación con seis niveles, estado fisiológico con dos niveles y línea con tres niveles),  $k$  es el número total de SNPs o loci,  $\mathbf{z}_j$  es un vector que representa el número de alelos del alelo de referencia para cada SNP o locus  $j$  (0, 1, ó 2),  $\alpha_j$  es el efecto de sustitución aleatorio para el locus  $j$ ,  $\delta_j$  es una variable aleatoria de 0 / 1 indicando la ausencia (con probabilidad  $\pi$ ) o presencia (con probabilidad  $1 - \pi$ ) del locus en el modelo. Los efectos  $\alpha_j$  tienen una distribución normal  $N(0, \sigma_{\alpha_j}^2)$ , y  $\mathbf{e}$  es un vector que representa los efectos residuales con una distribución normal  $N(0, \sigma_e^2)$  (Cesar et al., 2014; Meuwissen et al., 2001).

Las varianzas genéticas utilizadas a priori para el análisis con el método de Bayes B fueron calculadas por inferencia bayesiana con la base de datos de la selección divergente. Las varianzas y los  $\pi$  se muestran en la Tabla 3.

Tabla 3. Las varianzas y el valor pi ( $\pi$ ) para el método de Bayes B.

Carácter	Vg <sup>1</sup>	Ve <sup>2</sup>	$\pi$	SNP/Iter*
Tamaño de Camada	0.649500	5.255400	0.99947	60.8
Número de Nacidos Vivos	0.858864	9.819760	0.99947	60.8
Número de Nacidos Muertos	0.126133	0.665203	0.99947	60.8
Tasa de Ovulación	1.591300	3.381600	0.99954	52.7
Embriones Implantados	1.663800	5.898700	0.99954	52.7
Supervivencia Embrionaria	0.001156	0.027744	0.99955	51.6
Supervivencia Fetal	0.000896	0.005524	0.99954	52.7
Supervivencia Prenatal	0.000225	0.002275	0.99954	52.7

Vg<sup>1</sup>: varianza genética aditiva; Ve<sup>2</sup>: varianza residual; SNP/Iter<sup>3</sup>: número promedio de SNPs con efecto en cada iteración.

El número de SNPs con efecto diferente de cero en cada iteración es determinado utilizando un parámetro  $\pi$  (Tabla 3). Este parámetro fue calculado para cada carácter en función del número de datos y de SNPs (Ros-Freixedes et al., 2016). La fórmula utilizada fue:

$$\pi = 1 - \frac{\text{N}^{\circ} \text{ Datos}}{3 * (\text{N}^{\circ} \text{ SNPs})}$$

Los parámetros del modelo fueron estimados mediante funciones de distribuciones marginales posteriores. Estas distribuciones fueron estimadas para calcular todos los parámetros usando MCMC (Markov Chain Monte Carlos) (Fernando y Garrick, 2013). Después de varios análisis exploratorios, se utilizó un procedimiento en el cual se realizó un burn-in de 100,000 iteraciones, seguidas de 700,000 iteraciones. De estas iteraciones fueron utilizadas una de cada 60 iteraciones para estimar los efectos de los SNPs. Este procedimiento uso el software GenSel para obtener la distribución a posteriori de los efectos de los SNPs.

La inferencia de asociación en este modelo está basada en una ventana de una megabase (Cesar et al., 2014). Las ventanas genéticas fueron construidas a partir de los cromosomas y la distancia en pares de bases anotadas en el mapa de marcadores de *OryCun2.0* ensamblado.

En el presente estudio se utilizaron 2,171 ventanas asignadas a los 21 cromosomas autosómicos y al cromosoma X. La proporción de la varianza genética explicada por cada

ventana en una iteración en particular fue obtenida por la división entre la varianza de la ventana y la varianza del genoma entero en cada iteración. La ventana fue calculada con un modelo jerarquizado a partir del conjunto de SNPs que forma cada ventana. Para cada SNP consecutivo en una ventana se multiplicó el número de alelos del alelo de referencia de los SNPs por su efecto de sustitución en esa iteración. Además, se asume como una ventana relevante o QTL putativo, a aquella que explique más del 1% de la varianza genómica para cada carácter.

La posición de los genes candidatos fue investigada para cada ventana de una megabase (1-Mb) usando “UCSC Rabbit Genome Browser and Genome Browser” (Rosenbloom et al., 2015) y “Gen Ontology” (GO) (Ashburner et al., 2000). Se recuperaron las anotaciones de los genes de la base de datos “Ensembl Genes 85 Database” usando “Biomart Software” (Aken et al., 2016). La clasificación funcional y biológica de los genes fue realizada usando bases de datos en línea (“online”) sobre anotaciones genómicas: “GeneCards” (<http://www.genecards.org/>) y Base de Datos para Anotaciones, Visualización y Descubrimiento Integrado - DAVID (<http://David.abcc.ncifcrf.gov/>) a través de la herramienta bioinformática “Better Bunny” (Stelzer et al. 2016; Jiao et al. 2012; Craig et al. 2012).

Además, se realizó una evaluación del desequilibrio de ligamiento entre los SNPs de las ventanas asociados a varios caracteres y que se encontraran adyacentes en el mismo cromosoma. Estas fueron presentadas en un gráfico de bloques de desequilibrio de ligamiento que se generó con el paquete de R LDheatmap (Shin et al., 2006).

### **Estimación del Índice de Diferenciación Genética (F<sub>st</sub>).**

El índice de diferenciación genética mide la variación de las frecuencias alélicas entre dos líneas. Las líneas evaluadas fueron tres: línea alta de capacidad uterina (Alta), línea baja de capacidad uterina (Baja), y línea de referencia (línea V). Para el cálculo de este índice se utilizó el procedimiento de Weir y Cockerham (1984), en el cual se calcula para cada SNP:

$$F_{st} = \frac{\sigma_p^2}{\bar{p} (1 - \bar{p})}$$

En donde  $\bar{p}$  es la media de la frecuencia alélica para las dos líneas evaluadas y  $\sigma_p^2$  es la varianza de la frecuencia alélica entre las dos líneas. Esta varianza se estimó como:

$$\sigma_p^2 = (\overline{\rho^2}) - (\bar{\rho})^2$$

En donde  $(\overline{\rho^2})$  es la media del cuadrado de la frecuencia alélica en las dos líneas (Grams et al., 2015). Por otra parte, el término de  $F_{st}$  también puede ser definido como el índice de fijación o reducción de la heterocigosidad en subpoblaciones debido a la deriva génica. Por lo tanto podría ser calculada como un cociente de la sustracción entre la heterocigosidad esperada menos la observada, dividida entre la heterocigosidad observada.

Se consideraron SNPs con un efecto relevante, a los que presentaron un valor de  $F_{st}$  igual o mayor de 0.40 (valor umbral). El umbral se calculó como seis desviaciones estándar desde la media del  $F_{st}$  de todos los SNPs. Además, se definió como un “clúster” relevante aquellos que contenían por lo menos 4 SNPs relevantes y una distancia máxima de 2 Mb entre cada SNP relevante y menor de 5 Mb entre el primer y último SNP relevante del “clúster”. Como la selección produce un incremento de los índices  $F_{st}$  para una serie consecutiva de SNPs, se consideró que un clúster de SNPs relevantes daría más evidencias de una huella de selección (Carneiro et al., 2014b; Grams et al., 2015; Johansson et al., 2010)

La búsqueda y descripción de los genes dentro de estos “clúster” se realizó con el procedimiento descrito en el apartado anterior para el GWAS.

## Resultados

### Estudio de asociación para el tamaño de camada y sus componentes

El análisis de componentes principales representado en un gráfico de escalado multidimensional (MDS), muestra tres grupos de puntos que se corresponden con las dos líneas divergentes y la línea de referencia (Figura 1). Este resultado apoya la inclusión del efecto línea en el modelo estadístico para evitar un posible sesgo de los resultados por estratificación de la población.

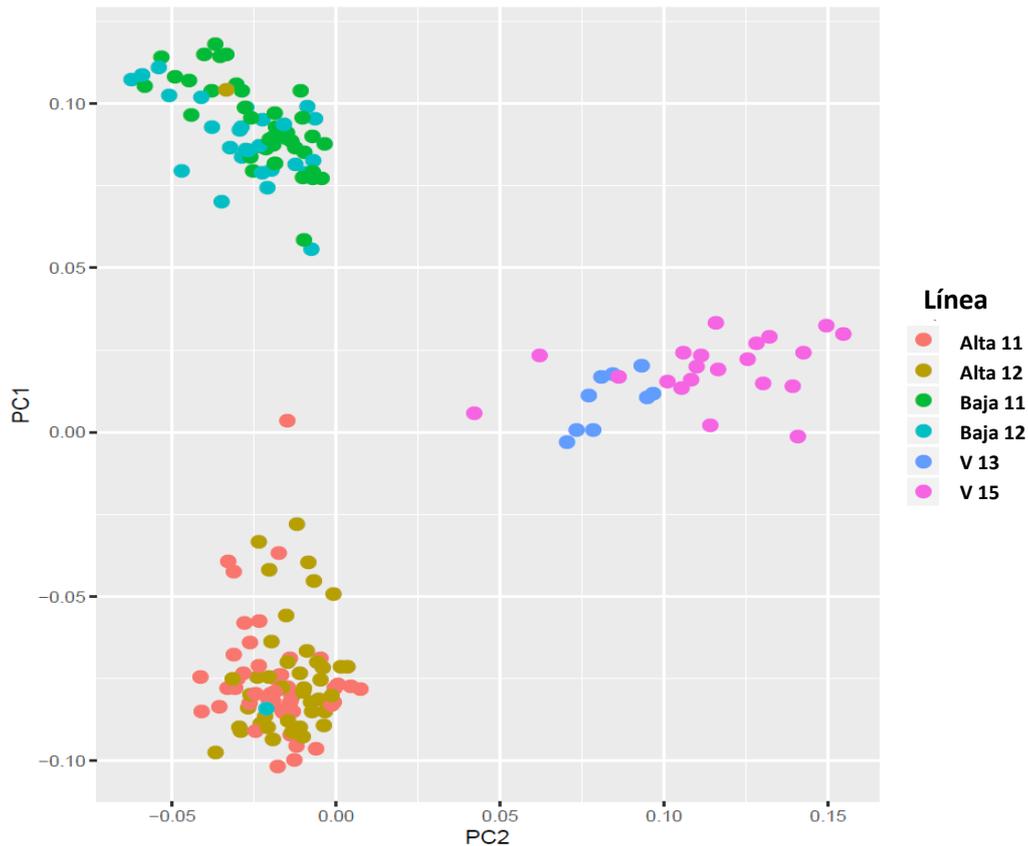


Figura 1. Gráfico de Escalado Multidimensional (MDS). Alta 11: Línea Alta de Capacidad Uterina de la Generación 11, Alta 12: Línea Alta de Capacidad Uterina de la Generación 12, Baja 11: Línea Baja de Capacidad Uterina de la Generación 11, Baja 12: Línea Baja de Capacidad Uterina de la generación 12, V 13: Línea V de la generación 13, V 15: Línea V de la generación 15. Los puntos se concentran en tres grandes grupos (líneas): Alta, Baja y V.

Los resultados de los análisis de asociación con el método Bayes B para los caracteres reproductivos se presentan a continuación.

### Tamaño de Camada y Número de Nacidos Vivos

La información de los resultados para el carácter tamaño de camada (LS, número de nacidos totales) y números de nacidos vivos (NBA) sobre las ventanas relevantes encontradas, el porcentaje de varianza genética explicado por el QTL putativo y los genes en esas regiones se muestran en la Tabla 4.

Tabla 4. Regiones de QTL asociadas con el tamaño de camada y número de nacidos vivos en conejo por Bayes B.

Id W <sup>1</sup>	Cr <sup>2</sup>	Número de SNP/W <sup>1</sup>	%Var <sup>3</sup>		Inicio/Final de posición en bp <sup>4</sup>	Genes candidatos
			Tamaño de Camada (LS)	Número de Nacido Vivos (NBA)		
1904	17	65	9.75	1.09	72,002,622-72,990,158	PNRC2, BMP4
1905	17	15	9.72	1.65	73,098,121-73,280,319	CDKN3, GMFB, CGRRF1, SAMD4A
2069	21	53	1.23	2.15	3,126,164-3,988,188	PES1, TCN2, SLC35E4, C5orf52, OSBP2, MORC2, SMTN, INPP5J, PLA2G3, RNF185, LIMK2, PIK3IP1, PATZ1, DRG1, EIF4ENIF1, SFI1
1903	17	50	1.15	1.11	71,014,965-71,994,214	PTGDR, PTGER2, TXNDC16, GPR137C, ERO1A, PSMC6, STYX, FERMT2, DDHD1

<sup>1</sup> W: ventana; <sup>2</sup> Cr: cromosoma; <sup>3</sup> %Var: porcentaje de varianza genética. <sup>4</sup> bp: pares de bases

Para LS se han localizado cuatro ventanas relevantes que representan el 21.84 % de la varianza genética y el 3.55% de la varianza fenotípica. Los QTL del cromosoma 17 representan el 94% del total de ventanas que explican la varianza genética de LS.

Para NBA, que es un carácter muy correlacionado genéticamente con LS, se encontraron las mismas regiones de QTLs asociadas. Estas ventanas representan el 6.0% de la varianza genética y el 0.64% de la varianza fenotípica de NBA. De acuerdo con lo observado para LS, un porcentaje elevado de la varianza genética (64%) de las ventanas asociadas, se debe a los QTL del cromosoma 17.

### Tasa de Ovulación

La única ventana relevante detectada para la tasa de ovulación representa un porcentaje muy bajo y no relevante (0.18%) de la varianza fenotípica (Tabla 5). En esta región sólo se identificó un gen codificante que se ha propuesto como gen candidato. Además, en esta región están localizados otros genes no codificantes, como el gen 5S-rRNA (gen ribosomal) y un gen pequeño nuclear snRNA denominado U6. Estos genes no codificantes no presentan anotaciones funcionales, biológicas o términos GO.

Tabla 5. Regiones de QTL asociadas con la tasa de ovulación en conejo por Bayes B.

Id W <sup>1</sup>	Cr <sup>2</sup>	Número de SNP/W <sup>1</sup>	%Var <sup>3</sup>	Inicio/Final de posición en bp <sup>4</sup>	Genes candidatos
1529	14	40	1.01	46,027,616- 46,968,731	MBNL1

<sup>1</sup> W: ventana; <sup>2</sup> Cr: cromosoma; <sup>3</sup> %Var: porcentaje de varianza genética. <sup>4</sup> bp: pares de bases

### Número de Embriones Implantados

Se han encontrado cinco regiones asociadas al número de embriones implantados que representan el 23.04 % de la varianza genética y el 4.95% de la varianza fenotípica (Tabla 6). El 64% de la varianza genética explicada por estas ventanas corresponde a los QTL del cromosoma 17.

Tabla 6. Regiones de QTL asociadas con el número embriones implantados en conejo por Bayes B.

Id W <sup>1</sup>	Cr <sup>2</sup>	Número de SNP/W <sup>1</sup>	%Var <sup>3</sup>	Inicio/Final de posición en bp <sup>4</sup>	Genes candidatos
1904	17	65	8.73	72,002,622- 72,990,158	PNRC2, BMP4
1905	17	15	6.03	73,098,121- 73,280,319	CDKN3, GMFB, CGRRF1, SAMD4A
1144	11	83	3.52	37,007,378- 37,982,329	---
1146	11	70	3.01	39,025,425- 39,997,550	TRIO, FAM105A, OTULIN, ANKH
1145	11	66	1.75	38,003,264- 38,987,355	DNAH5

<sup>1</sup> W: ventana; <sup>2</sup> Cr: cromosoma; <sup>3</sup> %Var: porcentaje de varianza, <sup>4</sup> bp: pares de bases.

Para el número de nacidos muertos (NBD) y las supervivencias embrionaria (ES), fetal (FS) y prenatal (PS) no se encontró ninguna ventana que explicara más de un 1% de la varianza genética por el método de Bayes B. En efecto, las ventanas con mayor valor para el porcentaje de la varianza genética explicada fue de 0.48% (Cr. 8) para NBD, 0.49% (Cr. 11) para ES, 0.32% (Cr. 2) para FS y 0.21% (Cr. 14) para PS.

Un estudio más detallado de las ventanas 1903, 1904 y 1905, que están asociadas a tres de los caracteres estudiados (LS, NBA, IE) revela que existen varios bloques de desequilibrio de ligamiento en esta región del cromosoma 17. La Figura 2 generada por LDheatmap (Shin et al., 2006) muestra los gráficos de desequilibrio de ligamiento (LD) en estas ventanas. A mayor valor de  $r^2$ , mayor LD y un color rojo más intenso. En la gráfica, los bloques de desequilibrio de ligamiento se observan como triángulos de color rojo, siendo la hipotenusa del triángulo la secuencia de SNPs del bloque de LD. Esto indica que el proceso de selección para capacidad uterina en la línea alta condujo a un incremento de los bloques de desequilibrio de ligamiento en esta región que se extiende por varios cientos de kilobases, siendo mucho menor el desequilibrio de ligamiento en la línea baja. Utilizando todos los datos aún permanece un bloque grande de 20 SNPs consecutivos con un  $r^2$  mayor de 0.8 (parte superior izquierda del gráfico).

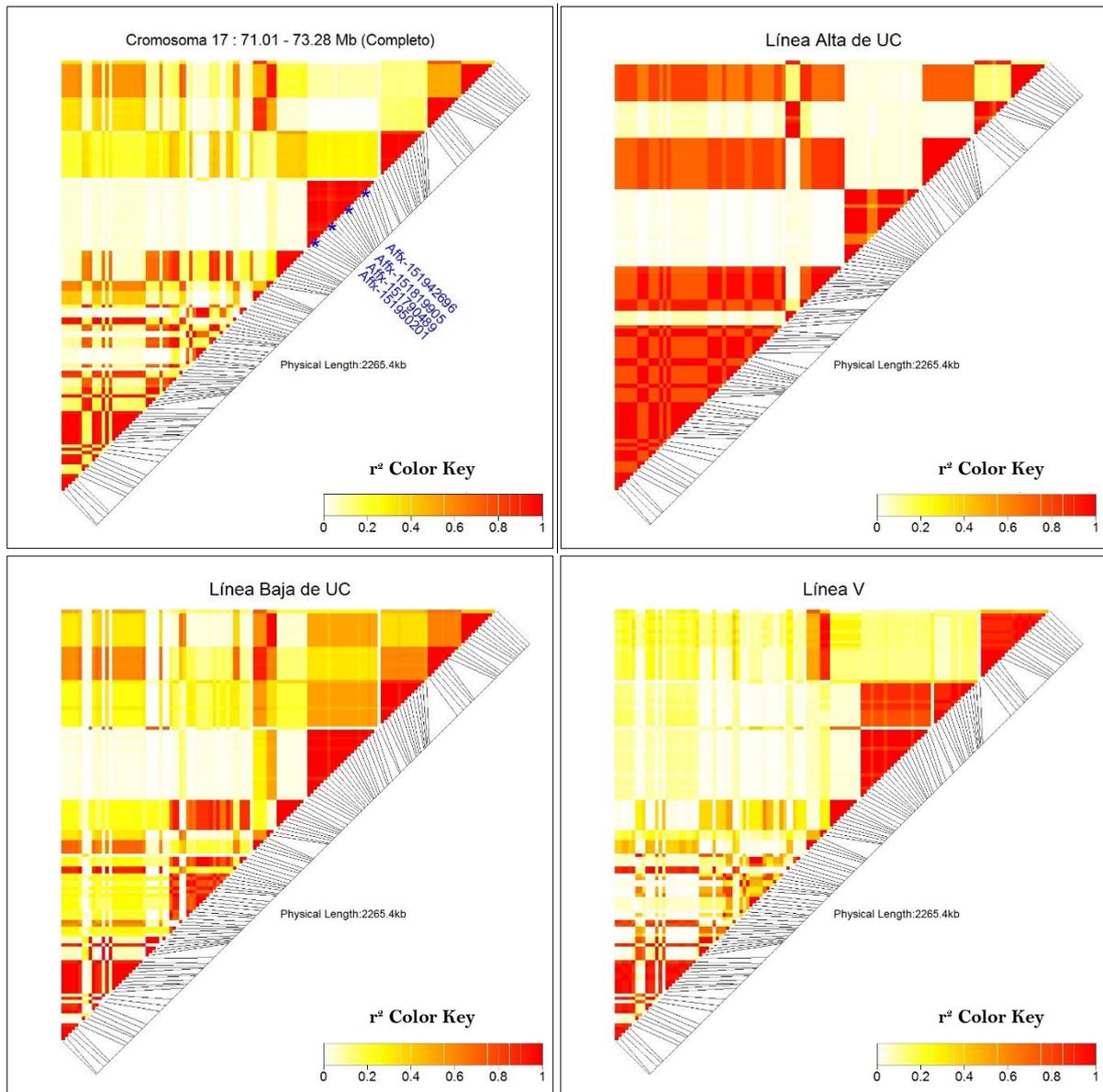


Figura 2. Desequilibrio de ligamiento entre los SNPs de las ventanas relevantes del Cromosoma17 para el conjunto de las líneas y para las líneas divergentes seleccionadas por alta y baja capacidad uterina (UC) y la línea de referencia V por separado. El primer SNP se encuentra en el vértice inferior de cada gráfica cuya posición es 71.01 Mb del cromosoma 17. El último SNP se encuentra en el vértice superior derecho de cada gráfica cuya posición es 73.28 Mb del cromosoma 17. Los bloques de disequilibrio de ligamiento se observan como triángulos rojos ( $r^2 \geq 0.8$ ).

## Diferenciación genética entre líneas

Para estimar los índices de diferenciación genética ( $F_{st}$ ) para las tres líneas estudiadas (Alta, Baja y V), se agruparon los animales de las dos generaciones y se consideró un solo grupo por línea.

Cuando se utilizaron las dos líneas divergentes para calcular los  $F_{st}$ , los valores de los índices de diferenciación genética indicaron una ausencia de huella genómica entre las líneas baja y alta, de acuerdo al umbral ( $\geq 0.40$ ) de nuestro estudio. Solo tres SNPs fueron relevantes ya que tenían valores de  $F_{st}$  mayores a 0.40 pero se encontraban en cromosomas distintos (cromosomas 9 y 13). Sin embargo, sí se detectaron huellas genómicas por efecto de la selección divergente por capacidad uterina cuando se calcularon los  $F_{st}$  de las líneas Alta y V, y de las líneas Baja y V. En la Tabla 7 se muestra el resumen de los “clúster” que se han clasificado relevantes. Se consideraron “clúster” relevantes a los que contenían al menos 4 SNP relevantes y presentaban una distancia máxima de 2 Mb entre SNPs y menor de 5 Mb entre el primer y último SNP del “clúster”.

Las dos regiones con mayor número de SNPs relevantes se encuentran en el cromosoma 11 y 13 para los  $F_{st}$  calculados con la línea baja de capacidad uterina y la línea de referencia (Tabla 7). En el Manhattan-plot de la Figura 3, se puede observar la cantidad de SNPs relevantes utilizando dos umbrales diferentes, uno mayor que 6 desviaciones estándar ( $F_{st} = 0.4$ ) y otro mayor que cuatro desviaciones estándar. El umbral utilizado,  $F_{st} = 0.4$ , genera menor cantidad de SNP relevantes, lo que podría explicar que el número de “clúster” relevantes también sea pequeño. Además, encontramos un “clúster” relevante en el cromosoma siete que se repite tanto para los  $F_{st}$  de la línea alta con la línea de referencia y la línea baja con la línea de referencia.

Cuando se intentó identificar genes en esas áreas genómicas utilizando el software Biomart, se observó que las regiones de los “clúster” relevantes encontrados para los  $F_{st}$  entre la línea alta y la de referencia, no presentaban anotaciones genómicas. Inclusive, el “clúster” 2 del cromosoma 7 (Tabla 7), descubierta en ambas comparaciones, es una región inter-génica; por lo tanto su alto  $F_{st}$  posiblemente podría ser el resultado de la deriva génica en las primeras generaciones de selección.

Tabla 7. Número y características de los “clúster” obtenidos a partir de los índices de diferenciación genética ( $F_{st}$ ) entre las poblaciones estudiadas, la línea alta de capacidad uterina (Alta) - la línea de referencia (línea V) y la línea baja de capacidad uterina (Baja) y la línea de referencia (línea V)

Líneas	Número de “clúster” <sup>1</sup>	Cr <sup>2</sup>	Inicio/Final de Posición en bp <sup>3</sup>	Largo en Mb	Número de SNPs relevantes
Alta UC* - V	1	7	3,622,155-3,680,724	0.06	5
	2*	7	167,905,624-167,990,601	0.08	4
	3	11	48,575,775-48,960,219	0.38	15
Baja UC* - V	4	1	152,437,729-153,032,499	0.59	4
	2*	7	167,905,624-167,990,601	0.08	4
	5	9	100,656,849-102,198,102	1.54	4
	6	11	27,415,436-29,341,239	1.93	26
	7	13	39,847,161-41,879,148	2.03	16
	8	X	77,441,343-77,784,401	0.34	4

<sup>1</sup> Un “clúster” se considera relevante cuando contenía al menos 4 SNP relevantes y presentaban una distancia máxima de 2 Mb entre SNPs y menor de 5 Mb entre el primer y último SNP del “clúster”. <sup>2</sup> Cr: cromosoma; <sup>3</sup> bp: pares de bases. \* es la misma región cromosómica para las dos comparaciones de líneas.

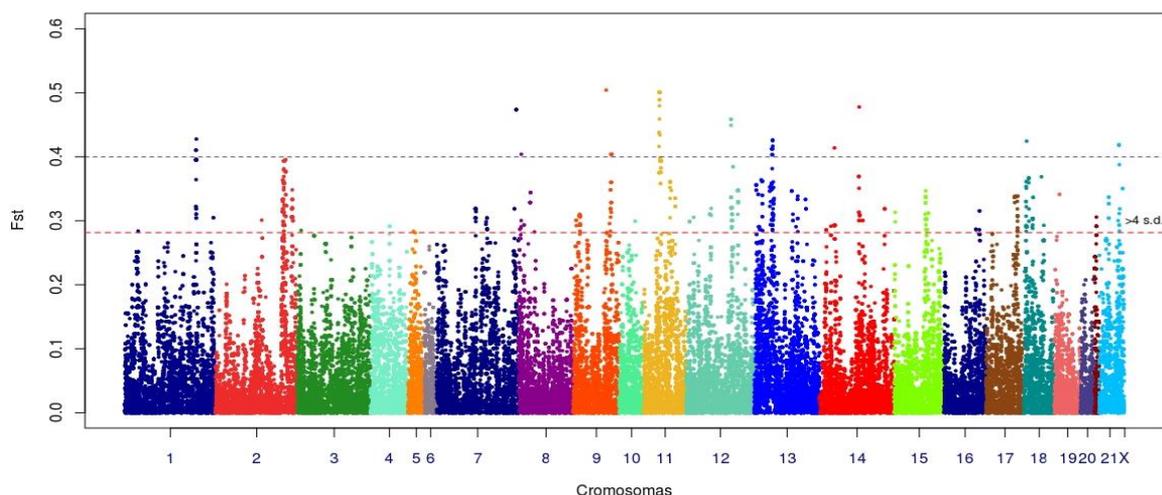


Figura 3. Manhattan-Plot de los valores  $F_{st}$  obtenidos de la línea baja de UC con la línea de referencia (línea V). Cada punto representa un  $F_{st}$ , y los diferentes colores representan los 22 pares de cromosomas del conejo. La línea discontinua de color naranja representa un umbral que se calculó como la media más cuatro desviaciones típicas. La línea discontinua negra representa el umbral para un valor de  $F_{st}$  igual a 0.40.

El análisis para identificar genes funcionales para los “clúster” que presentaron índices de diferenciación génicas relevantes entre la línea baja de UC y la línea de referencia, muestra que existen una gran cantidad de genes candidatos por posición de acuerdo al tamaño del “clúster” (Tabla 8). En efecto, el cromosoma 13 tiene 53 genes anotados, estos genes podrían estar involucrados en las huellas de selección después de la selección por baja capacidad uterina; de todos ellos solo 22 han sido propuestos como genes candidatos. Las regiones del autosoma 13 y el cromosoma sexual “X” son los únicos que presentan genes involucrados directamente con procesos reproductivos como la espermatogénesis, la fecundación, el desarrollo placentario, el desarrollo ovárico y el sistema endocrino femenino. Estos genes son: “Tudor and KH Domain Containing” (TDRKH), Ornithine Decarboxylase Antizyme 3 (OAZ3), “Trinucleotide Repeat-Containing Gene 4 Protein” (TNRC4), “Aryl Hydrocarbon Receptor Nuclear Translocator” (ARNT) y Dachshund family Transcription factor 2 (DACH2). Además hay otro conjunto de genes relacionados con procesos de regulación negativa en la respuesta inmunológica (células T), la mineralización de huesos y el proceso de apoptosis (proliferación celular), que podrían haber contribuido a la disminución de tamaño de camada en la línea baja. Entre ellos tenemos los genes: “FER Tyrosine Kinase” (FER), “Myeloid Cell Leukemia 1” (MCL1), “Extracelular Matrix Protein 1” (ECM1), “BCL 2 / Adenovirus E1B 19 KD Indirecting Protein Like” (BNIPL), “Proteasome subunit beta Type – 4” (PSMB4), y “Tumor Necrosis Factor – Alpha induced protein 8 – like 2” (TNFAIP8L2).

Tabla 8. Número de genes candidatos, principales genes candidatos y funciones biológicas de los “clúster” obtenidos a partir de los índices de diferenciación genética (Fst) entre la línea baja de capacidad uterina (Baja) y la línea de referencia (línea V).

Cr <sup>1</sup>	Número de GC <sup>2</sup>	Principales Genes Candidatos	Funciones Biológicas
1	3	CTR9, EIF4G2, GALNT18	- Asociado a RNA Polimerasa y el proceso de transcripción. - Glicosilación Proteica
9	3	MC4R, CDH20, RNF152	- Receptores transmembranas y activación de vías de señales - Regulación de procesos metabólicos - Unión de iones metálicos - Lípidos de membrana
11	4	FBXL17, FER, PJA2, MAN2A1	- Regulación de señales de factores de crecimiento epidérmico - Proliferación celular y fosforilación. - Procesos metabólicos de manosa. - Actividad catalítica y desarrollo del hígado. - Regulación de Proteína Quinasa A - Unión de iones metálicos (Zinc)
13	53	RPTN, S100A11, THEM4, THEM5, RORC, TDKRKH, OAZ3, MRPL9, TNRC4, SNX27, POGZ, PSMB4, ARNT, SELENBP1, MLLT11, BNIPL, ANXA9, ARNT, MCL1, CTSS, ECM1, TNFAIP8L2	- Unión de iones metálicos (selenio, calcio, zinc). - Actividad de hidrolisis palmitoil-CoA - Metabolismo de ácidos grasos - Rutas señales mediadas por esteroides - Espermatogénesis, metilación en gametos, fertilidad y meiosis en machos - Desarrollo Embrionario de la Placenta - Constituyente estructural en ribosomas - Unión de nucleótidos y proteínas - Regulación de splicing de mRNA. - Procesos catabólicos celulares-proteólisis - Actividad metabólica fosfatilinositol. - Regulación negativa de respuesta inflamatoria (antígeno) y activación de células T. - Regulación de factores de transcripción - Respuesta a acidez de pH. - Regulación Negativa de Mineralización de huesos. - Regulación de tasa de crecimiento y proliferación celular.
X	1	DACH2	- Desarrollo de primeras características sexuales en la hembra. - Desarrollo ovárico y respuesta de secreciones hormonales.

Un “clúster” se considera relevante cuando contenía al menos 4 SNPs relevantes y presentaban una distancia máxima de 2 Mb entre SNPs y menor de 5 Mb entre el primer y último SNP del “clúster”. Cr<sup>1</sup>: cromosoma, GC<sup>2</sup>: genes candidatos.

En la [Tabla S1](#) del Material Suplementario se encuentra las abreviaturas y nombres de los genes presentes en este documento.

## Discusión

El número de ventanas para tamaño de camada en nuestro estudio es pequeño comparado con los estudios publicados en porcino ([Onteru et al., 2011](#); [Schneider et al., 2014](#)). Las ventanas encontradas para el tamaño de camada y el número de nacidos vivos son las mismas, lo cual puede ser explicado por la alta correlación genética entre ambos caracteres en conejo ([Blasco et al., 2005](#); [García y Baselga, 2002b](#)).

Las anotaciones en las bases de datos de conejo están incompletas en cuanto a la información de evidencia funcional sobre los genes de la especie. Por este motivo, para comprender la función de los genes y su participación en las diferentes rutas biológicas se utilizó las anotaciones de los ortólogos en humano, ratón y rata. De acuerdo a los resultados de las anotaciones obtenidas para el tamaño de camada y el número de nacidos vivos, existe un “clúster” de genes que presentan una gran asociación por sus anotaciones biológicas: PTGDR (Prostagaldin D<sub>2</sub> Receptor), PTGER2 (“Prostaglandin E Receptor 2”), PIK3IP1 (phosphoinositide-3-kinase interacting protein 1) y GPR137C (G protein-coupled receptor 137C). Una de las principales funciones biológicas de este grupo de genes es la actividad de receptores icosanoides y prostanoides (prostaglandinas), función atribuida principalmente a los genes receptores de la prostaglandinas (PTGDR y PTGER2). Las anotaciones de otro grupo de genes como son los genes BMP4, LIMK2, PTGDR y STYX, están principalmente asociadas a procesos reproductivos en humanos, rata y ratón. Otras anotaciones funcionales de interés son las relacionadas con las moléculas asociadas al citoplasma (en la que participarían 12 genes, entre ellos PNRC2, CDKN3, LIMK2, SAMD4A) y la regulación positiva de procesos metabólicos (cuatro genes que son BMP4, PSMC6 y SAMD4A).

Las prostaglandinas son necesarias para el buen funcionamiento del endometrio, miometrio y útero. Los genes de los receptores de prostaglandina F (PTGFR) tienen un papel más relevante durante el ciclo reproductivo de la hembra y la embriogénesis que los otros genes receptores de la prostaglandinas ([Blesson y Sahlin, 2014](#)). Por otra parte, el gen receptor de la prostaglandina PTGDR tiene mayor importancia en el macho para la señalización de la diferenciación de células de Sertoli y germinales en el desarrollo embrionario de los testículo y la espermatogénesis en adultos ([Rossitto et al., 2014](#)). No obstante, una alteración en el

reconocimiento de la prostaglandina tendría efectos sobre la expansión del *cumulus* que rodea al ovocito y problemas en el parto en la hembra (Matzuk y Lamb, 2002).

Por otro lado, los genes de la familia factor-beta de crecimiento y transformación como el gen BMP4 están ampliamente relacionados con el crecimiento y desarrollo folicular en los ovarios de mamíferos (Al-Samerria et al., 2015; Zhu et al., 2013). Estos genes son expresados en las células de la teca y junto con el gen IGF previenen o retrasan la atresia o luteinización en bovino y caprino, modulan la esteroidogénesis y la regulación de los receptores de las hormonas folículo estimulante y luteinizante para incrementar el número de células viables (Glistler et al., 2004; Zhu et al., 2013). Por lo tanto, BMP4 es esencial para el desarrollo de los folículos primordiales, su supervivencia y su potencial calidad para los subsecuentes procesos embrionarios, ya que previene la apoptosis de oocitos por la regulación de *Sohlh2* y *c-kit* (Ding et al., 2013). No obstante, en nuestro estudio este gen no fue identificado como una región asociada a la tasa de ovulación sino al tamaño de camada. Respecto a su posible papel en el tamaño de camada, el gen BMP4 en humanos se ha visto que está involucrado con el desarrollo del trofoblasto, que es un tejido que rodea al blastocisto y que está implicado en la fijación e implantación del embrión y en la posterior formación de la placenta (Li y Parast, 2014). Por otra parte, los genes LIMK2 y STYX influyen en la degeneración de espermatogonias en los túbulos seminíferos, incrementa la apoptosis y defectos en la desarrollo de las espermátidas (Matzuk y Lamb, 2002).

En cuanto a los genes de mayor relevancia asociados con el carácter número de embriones implantados tenemos el BMP4, ANKH, CGRRF11, CDKN3 y SAMD4A. Estos genes están involucrados en funciones de regulación negativa de la proliferación celular, la regularización de la mineralización de los huesos y la formación biomineral. Además, en ratón, los genes BMP4, ANKH y CDKN3 parecen estar involucrados en procesos de desarrollo embrionario. (Goggolidou et al., 2013; Gurley et al., 2006; Lawson et al., 1999)

De todos los genes encontrados en esta regiones genómicas, únicamente el gen ERO1A precursor del transcrito ER1L, se ha encontrado asociado con diferencias de expresión en las líneas de capacidad uterina en conejos (Ballester et al., 2013). El gen ERO1A está relacionado con el mantenimiento de la homeostasis dentro del retículo endoplasmático y una sobreexpresión puede generar apoptosis en las células (Benham et al., 2013; Konno et al., 2015). Sin embargo, no se encontraron regiones asociadas en los cromosomas 1,13 y X, que es donde se localizan tres de los genes candidatos propuestos en previos artículos para la

capacidad uterina en conejo. Estos genes propuestos son el gen PGR (receptor de progesterona), el gen OVGP1 (glicoproteína oviductal 1) y el gen TIMP1 (tejido inhibidor de la metaloproteínasa 1), respectivamente (Argente et al., 2010; Merchán et al., 2009; Peiro et al., 2010).

Los genes encontrados en nuestro análisis de asociación no se corresponden con los encontrados en porcino para la tasa de ovulación (Schneider et al., 2014), el tamaño de camada y el número de nacidos vivos (Onteru et al., 2011; Schneider et al., 2012; Spötter y Distl, 2006). De acuerdo con lo publicado con otros autores, se han encontrado QTLs asociados a varios caracteres. En nuestro estudio, dos QTL en el cromosoma 17 se solapan para LS, NBA y IE. Schneider et al. (2012) encontraron varias regiones de QTL solapadas para el tamaño de camada, el número de nacidos vivos y muertos y el promedio del peso de la camada en una población de cerdos compuesta por reproductores de tres razas. Otros estudios realizados en porcino indican que los genes que se expresan para el tamaño de camada pueden ser diferentes a los expresados para otros componentes de tamaño de camada (Onteru et al., 2011; Schneider et al., 2014). Además, los genes expresados podrían ser diferentes dependiendo del orden del parto y de la población estudiada; esto dificultaría utilizar esta información en otras poblaciones comerciales (Blasco y Toro, 2014; Onteru et al., 2011; Sell-Kubiak et al., 2015). Bergfelder-Drüing et al. (2015) no encontraron QTLs solapados para número de nacidos vivos en diferentes poblaciones comerciales de cerdo utilizando varios partos y cuestionan la utilidad de las estrategias de poblaciones divergentes para detectar QTL relevantes o mejorar la precisión de la selección genómica.

En el análisis de los índices de diferenciación genética ( $F_{st}$ ) no se detectaron huellas de selección para las líneas alta y baja. Estos resultados están en desacuerdo con los resultados publicados por Grams et al. (2015). En ese trabajo se encontraron 13 “clúster” en dos poblaciones divergentes de gallinas ponedoras para el carácter comportamiento agresivo. En ese estudio la media de  $F_{st}$  entre las dos líneas fue de 0.15 versus el encontrado en las líneas divergentes por capacidad uterina que fue de 0.053. No obstante, la media de 0.053 es superior a las encontradas al estudiar el  $F_{st}$  de la línea de referencia con cada una de las líneas divergentes (0.046 en ambos casos), en las que sí se detectaron huellas genómicas por efecto de la selección divergente por capacidad uterina. Una posible explicación es que la selección en las líneas alta y baja haya modificado la frecuencia de diferentes genes en cada una de las líneas divergentes, así habría más genes afectados, pero la magnitud del cambio de frecuencias de esos genes sería menor y más difícil de detectar. En los estudios realizados comparando

conejo doméstico y salvaje, se han detectado hasta 78 regiones con  $F_{st} > 0.35$  y una heterocigosidad mayor a 0.05, con un gran cantidad de estas regiones en el cromosoma X (Carneiro et al., 2014a). Sin embargo en subespecies y poblaciones comerciales de conejo, no se han observado estos valores tan altos de  $F_{st}$ , lo cual implica que el conejo presenta varios sitios de barridos selectivos en su genoma por la domesticación, y una disminución de la variabilidad en las poblaciones comerciales de conejo (Carneiro et al., 2014b). Estos resultados, están de acuerdo con la ausencia de detección de huellas genómicas en las líneas divergentes de capacidad uterina ya que provienen de una población cerrada con varias generaciones de selección por tamaño de camada.

En este trabajo sólo se encontraron seis “clúster” como huellas de selección para los índices de diferenciación genética entre la línea baja y V. Precisamente un solo “clúster” fue encontrado en el cromosoma X, el cual corresponde al gen *DACH2*. Los alelos favorables del gen *DACH2* deben estar en frecuencias altas en la línea V pero no así en la línea baja. La naturaleza de los resultados de  $F_{st}$  indica que la selección aumento varias regiones y que la interacción entre estas regiones podría ser la responsable de expresar la divergencia entre las líneas. Por lo tanto, podemos deducir que la magnitud de la capacidad uterina y tamaño de camada presentan una naturaleza poligénica (Carneiro et al., 2014a; Peripato et al., 2004) y que otros factores como la epistasia podrían haber influido en estos caracteres.

## **Conclusiones e Implicaciones**

Para concluir, las regiones genómicas y los genes identificados por este estudio muestran un avance en el conocimiento del tamaño de camada en conejo. Este estudio es el primer análisis de asociación de genoma completo en conejo y para ello se ha utilizado un panel de 200 K de Affymetrix. Los resultados revelan cuatro áreas importantes para el tamaño de camada y un posible efecto pleiotrópico de algunos genes del cromosoma 17 por el solapamiento de QTLs para varios caracteres (tamaño de camada al nacimiento, número de nacidos vivos y el número de embriones implantados). Unos genes candidatos fueron identificados por su plausible función en procesos reproductivos, entre ellos cabe destacar *BMP4* y *PTGDR*. Por otra parte, en el estudio de búsqueda de huellas de selección no se encontró ninguna huella de selección entre las líneas divergentes pero sí entre la línea seleccionada para disminuir la capacidad uterina y la línea de referencia. Estos resultados apoyan la hipótesis de que en la línea alta de capacidad uterina, los genes con efectos grandes ya están fijados o con frecuencias muy elevada, lo que es coherente con el hecho de que los animales de la población base de las líneas

divergentes provienen de una línea seleccionada durante 12 generaciones por tamaño de camada.

Nuestros hallazgos ofrecen un nuevo punto de partida para futuras investigaciones sobre el uso de otras estrategias de selección para mejorar el tamaño de camada o la aplicación de la selección genómica en poblaciones comerciales de conejo.

## Agradecimientos

El trabajo ha sido financiado por la Secretaría de Estado I+D+i del Ministerio de Economía y Competitividad de España (AGL2014-55921-C2-1-P), y B. S. Sosa-Madrid por la Secretaría Nacional de Ciencia, Innovación y Tecnología (SENACYT) de la República de Panamá.

## Referencias Bibliográficas

- Aken, B.L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez Banet, J., Billis, K., García Girón, C., Hourlier, T., et al. (2016). The Ensembl gene annotation system. *Database* 2016, baw093.
- Al-Samerria, S., Al-Ali, I., McFarlane, J.R., y Almahbobi, G. (2015). The impact of passive immunisation against BMPRII and BMP4 on follicle development and ovulation in mice. *Reproduction* 149, 403–411.
- Anderson, C. a, Pettersson, F.F.H., Clarke, G.G.M., Cardon, L.R., Morris, A.P., y Zondervan, K.T. (2010). Data quality control in genetic case-control association studies. *Nature protocols* 5, 1564–1573.
- Argente, M.J., Santacreu, M.A., Climent, A., Bolet, G., y Blasco, A. (1997). Divergent selection for uterine capacity in rabbits. *Journal of Animal Science* 75, 2350–2354.
- Argente, M.J., Santacreu, M.A., Climent, A., y Blasco, A. (2000). Genetic Correlations between Litter Size and Uterine Capacity. 7th World Rabbit Congress 8, 333–338.
- Argente, M.J., Merchán, M., Peiró, R., García, M.L., Santacreu, M. a., Folch, J.M., y Blasco, A. (2010). Candidate gene analysis for reproductive traits in two lines of rabbits divergently selected for uterine capacity. *Journal of Animal Science* 88, 828–836.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics* 25, 25–29.
- Ballester, M., Castelló, A., Peiró, R., Argente, M.J., Santacreu, M.A., y Folch, J.M. (2013). Identification of differentially expressed genes in the oviduct of two rabbit lines

- divergently selected for uterine capacity using suppression subtractive hybridization. *Animal genetics* 44, 296–304.
- Baselga, M., y Blasco, A. (1989). *Mejora Génética del Conejo de Producción de Carne* (Madrid, España: Ediciones Mundi-Prensa).
- Benham, A.M., van Lith, M., Sitia, R., y Braakman, I. (2013). Ero1-PDI interactions, the response to redox flux and the implications for disulfide bond formation in the mammalian endoplasmic reticulum. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, 20110403–20110403.
- Bergfelder-Drüing, S., Grosse-Brinkhaus, C., Lind, B., Erbe, M., Schellander, K., Simianer, H., y Tholen, E. (2015). A genome-wide association study in large white and landrace pig populations for number piglets born alive. *PLoS ONE* 10, 1–19.
- Blasco, A., y Toro, M.A. (2014). A short critical history of the application of genomics to animal breeding. *Livestock Science* 166, 4–9.
- Blasco, A., Argente, M.J., Haleyt, C.S., y Santacreu, M.A. (1994). Relationships Between Components of Litter Size in Unilaterally Ovariectomized and Intact Rabbit Does. *Journal of Animal Science* 72, 3066–3072.
- Blasco, A., Ortega, J.A., Climent, A., y Santacreu, M.A. (2005). Divergent Selection for Uterine Capacity in Rabbits . I . Genetic Parameters and Response to Selection. *Journal Animal Science* 83, 2297–2302.
- Blesson, C.S., y Sahlin, L. (2014). Prostaglandin E and F receptors in the uterus. *Receptors & Clinical Investigation* 2, 1–12.
- Browning, B.L., y Browning, S.R. (2009). A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *The American Journal of Human Genetics* 84, 210–223.
- Carneiro, M., Albert, F.W., Afonso, S., Pereira, R.J., Burbano, H., Campos, R., Melo-Ferreira, J., Blanco-Aguiar, J.A., Villafuerte, R., Nachman, M.W., et al. (2014a). The Genomic Architecture of Population Divergence between Subspecies of the European Rabbit. *PLoS Genetics* 10, e1003519.
- Carneiro, M., Rubin, C., Di Palma, F., Albert, F.W., Alfoldi, J., Barrio, A.M., Pielberg, G., Rafati, N., Sayyab, S., Turner-Maier, J., et al. (2014b). Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* 345, 1074–1079.
- Cesar, A.S., Regitano, L.C., Tullio, R.R., Lanna, D.P., Nassu, R.T., Mudado, M. a, Oliveira, P.S., do Nascimento, M.L., Chaves, A.S., Alencar, M.M., et al. (2014). Genome-wide association study for intramuscular fat deposition and composition in Nellore

- cattle. *BMC genetics* 15, 39.
- Christenson, R.K., Leymaster, K.A., y Young, L.D. (1987). Justification of unilateral hysterectomy-ovariectomy as a model to evaluate uterine capacity in swine. *Journal of animal science* 65, 738–744.
- Craig, D.B., Kannan, S., y Dombkowski, A. a (2012). Augmented annotation and orthologue analysis for *Oryctolagus cuniculus*: Better Bunny. *BMC Bioinformatics* 13, 84.
- Ding, X., Zhang, X., Mu, Y., Li, Y., y Hao, J. (2013). Effects of BMP4/SMAD signaling pathway on mouse primordial follicle growth and survival via up-regulation of *Sohlh2* and *c-kit*. *Molecular Reproduction and Development* 80, 70–78.
- Fernando, R.L., y Garrick, D. (2013). Bayesian methods applied to GWAS. In *Genome-Wide Association Studies and Genomic Prediction*, C. Gondro, J.H. van der Werf, y B. Hayes, eds. (Human Press Inc.), pp. 237–274.
- García, M.L., y Baselga, M. (2002a). Genetic Response to Selection for Reproductive Performance in a Maternal Line of Rabbits. *World Rabbit Science* 10, 71–76.
- García, M.L., y Baselga, M. (2002b). Estimation of Genetic Response to selection in litter size of rabbits using a cryopreserved control population. *Livestock Production Science* 74, 45–53.
- Glister, C., Kemp, C.F., y Knight, P.G. (2004). Bone morphogenetic protein (BMP) ligands and receptors in bovine ovarian follicle cells: actions of BMP-4, -6 and -7 on granulosa cells and differential modulation of Smad-1 phosphorylation by follistatin. *Reproduction* 127, 239–254.
- Gogolidou, P., Soneji, S., Powles-Glover, N., Williams, D., Sethi, S., Baban, D., Simon, M.M., Ragoussis, I., y Norris, D.P. (2013). A chronological expression profile of gene activity during embryonic mouse brain development. *Mammalian Genome* 24, 459–472.
- Grams, V., Wellmann, R., Preuß, S., Grashorn, M.A., Kjaer, J.B., Bessei, W., y Bennewitz, J. (2015). Genetic parameters and signatures of selection in two divergent laying hen lines selected for feather pecking behaviour. *Genetics Selection Evolution* 47, 77.
- Gurley, K.A., Chen, H., Guenther, C., Nguyen, E.T., Rountree, R.B., Schoor, M., y Kingsley, D.M. (2006). Mineral formation in joints caused by complete or joint-specific loss of ANK function. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* 21, 1238–1247.
- Jiao, X., Sherman, B.T., Huang, D.W., Stephens, R., Baseler, M.W., Lane, H.C., y Lempicki, R.A. (2012). DAVID-WS: A stateful web service to facilitate gene/protein list analysis. *Bioinformatics* 28, 1805–1806.

- Johansson, A.M., Pettersson, M.E., Siegel, P.B., y Carlborg, Ö. (2010). Genome-Wide Effects of Long-Term Divergent Selection. *PLoS Genetics* 6, e1001188.
- Khalil, M.H., y Al-Saef, A.M. (2008). Methods, criteria, techniques and genetic responses for rabbit selection: A review. *Proc. 9th World Rabbit Congress* 1–22.
- Konno, T., Pinho Melo, E., Lopes, C., Mehmeti, I., Lenzen, S., Ron, D., y Avezov, E. (2015). ERO1-independent production of H<sub>2</sub>O<sub>2</sub> within the endoplasmic reticulum fuels Prdx4-mediated oxidative protein folding. *The Journal of Cell Biology* 211, 253–259.
- Laborda, P., Mocé, M.L., Blasco, A., y Santacreu, M.A. (2012). Selection for ovulation rate in rabbits: Genetic parameters and correlated responses on survival rates. *Journal of Animal Science* 90, 439–446.
- Lawson, K.A., Dunn, N.R., Roelen, B.A.J., Lawson, K.A., Dunn, N.R., Roelen, B.A.J., Zeinstra, L.M., Davis, A.M., Wright, C.V.E., Korving, J.P.W.F.M., et al. (1999). mouse embryo Bmp4 is required for the generation of primordial germ cells in the mouse embryo. 424–436.
- Li, Y., y Parast, M.M. (2014). BMP4 regulation of human trophoblast development. *International Journal of Developmental Biology* 58, 239–246.
- Marras, G., Rossoni, A., Schwarzenbacher, H., Biffani, S., Biscarini, F., Stella, A., y Nicolazzi, E.L. (2015). ZANARDI an open-source pipeline for multiple-species genomic analysis using SNP array data.
- Matzuk, M.M., y Lamb, D.J. (2002). Genetic dissection of mammalian fertility pathways. *Nature Cell Biology* 4, S33–S40.
- Merchán, M., Peiró, R., Argente, M.J., Santacreu, M.A., García, M.L., Blasco, A., y Folch, J.M. (2009). Analysis of the oviductal glycoprotein 1 polymorphisms and their effects on components of litter size in rabbits. *Animal Genetics* 40, 756–758.
- Meuwissen, T.H.E., Hayes, B.J., y Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Mocé, M.L., Santacreu, M.A., Climent, A., y Blasco, A. (2004). The effect of divergent selection for uterine capacity on fetal and placental development at term in rabbits: Maternal and embryonic genetic effects. *Journal of Animal Science* 82, 1046–1052.
- Mocé, M.L., Santacreu, M.A., Climent, A., y Blasco, A. (2005). Divergent selection for uterine capacity in rabbits . III . Responses in uterine capacity and its components estimated with a cryopreserved control population. *Journal of Animal Science* 83, 2308–2312.
- Onteru, S.K., Fan, B., Du, Z.-Q., Garrick, D.J., Stalder, K.J., y Rothschild, M.F. (2011). A

- whole-genome association study for pig reproductive traits. *Animal Genetics* *43*, 18–26.
- Peiro, R., Herrler, A., Santacreu, M.A., Merchan, M., Argente, M.J., Garcia, M.L., Folch, J.M., y Blasco, A. (2010). Expression of progesterone receptor related to the polymorphism in the PGR gene in the rabbit reproductive tract. *Journal of Animal Science* *88*, 421–427.
- Peripato, A.C., De Brito, R.A., Matioli, S.R., Pletscher, L.S., Vaughn, T.T., y Cheverud, J.M. (2004). Epistasis affecting litter size in mice. *Journal of Evolutionary Biology* *17*, 593–602.
- Ragab, M., Sánchez, J.P., Mínguez, C., Vicente, J.S., y Baselga, M. (2014). Litter size components in a full diallel cross of four maternal lines of rabbits. *Journal Animal Science* *92*, 3231–3236.
- Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2015). The UCSC Genome Browser database: 2015 update. *Nucleic Acids Research* *43*, D670–D681.
- Ros-Freixedes, R., Gol, S., Pena, R.N., Tor, M., Ibáñez-Escriche, N., Dekkers, J.C.M., y Estany, J. (2016). Genome-Wide Association Study Singles Out SCD and LEPR as the Two Main Loci Influencing Intramuscular Fat Content and Fatty Acid Composition in Duroc Pigs. *PLOS ONE* *11*, e0152496.
- Rossitto, M., Ujjan, S., Poulat, F., y Boizet-Bonhoure, B. (2014). Multiple roles of the prostaglandin D<sub>2</sub> signaling pathway in reproduction. *Reproduction* *149*, R49–R58.
- Santacreu, M.A., Mocé, M.L., Climent, A., y Blasco, A. (2005). Divergent selection for uterine capacity in rabbits . II. Correlated Response in Litter Size and its components estimated with a cryopreserved control population. *Journal of Animal Science* *83*, 2303–2307.
- Schneider, J.F., Rempel, L. a., Snelling, W.M., Wiedmann, R.T., Nonneman, D.J., y Rohrer, G.A. (2012). Genome-wide association study of swine farrowing traits. Part II: Bayesian analysis of marker data., *Journal of Animal Science* *90*, 3360–3367.
- Schneider, J.F., Nonneman, D.J., Wiedmann, R.T., Vallet, J.L., y Rohrer, G.A. (2014). Genomewide association and identification of candidate genes for ovulation rate in swine. *Journal of Animal Science* *92*, 3792–3803.
- Sell-Kubiak, E., Duijvesteijn, N., Lopes, M.S., Janss, L.L.G., Knol, E.F., Bijma, P., y Mulder, H.A. (2015). Genome-wide association study reveals novel loci for litter size and its variability in a Large White pig population. *BMC Genomics* *16*, 1–13.

- Shin, J.-H., Blay, S., McNeney, B., y Graham, J. (2006). LDheatmap: An R Function for Graphical Display of Pairwise Linkage Disequilibria between Single Nucleotide Polymorphisms. *Journal of Statistical Software* 16, 1–10.
- Spötter, A., y Distl, O. (2006). Genetic approaches to the improvement of fertility traits in the pig. *The Veterinary Journal* 172, 234–247.
- Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, I., Mazor, Y., et al. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. In *Current Protocols in Bioinformatics*, (Hoboken, NJ, USA: John Wiley & Sons, Inc.), p. 1.30.1-1.30.33.
- Weir, B.S., y Cockerham, C.C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 38, 1358.
- Zhu, G., Cui, Y., Qinglin-Wang, Yonggang-Kang, Yanzhi-Lv, Wang, J., Song, Y., y Cao, B. (2013). Bone morphogenetic proteins (BMP) 2, 4, 6 and 7 affect ovarian follicular development through regulation of follicle-stimulating hormone receptor (FSHR) and luteinizing hormone receptor (LHR) expression in goat granulosa cells. *Journal of Cell Biology and Genetics* 3, 14–21.
- Ziadi, C., Moce, M.L., Laborda, P., Blasco, A., y Santacreu, M. a (2013). Genetic selection for ovulation rate and litter size in rabbits: Estimation of genetic parameters and direct and correlated responses. *Journal of Animal Science* 91, 3113–3120.

**Tabla S1**

Chromosome Name	Associated Gene Name	Description
1	CTR9	CTR9 homolog, Paf1/RNA polymerase II complex component
1	EIF4G2	eukaryotic translation initiation factor 4 gamma 2
1	GALNT18	polypeptide N-acetylgalactosaminyltransferase 18
9	MC4R	melanocortin 4 receptor
9	CDH20	cadherin 20
9	RNF152	ring finger protein 152
11	TRIO	trio Rho guanine nucleotide exchange factor
11	FAM105A	family with sequence similarity 105 member A
11	OTULIN	OTU deubiquitinase with linear linkage specificity
11	ANKH	ANKH inorganic pyrophosphate transport regulator
11	DNAH5	dynein axonemal heavy chain 5
11	FBXL17	F-box and leucine rich repeat protein 17
11	FER	FER tyrosine kinase
11	PJA2	praja ring finger ubiquitin ligase 2
11	MAN2A1	mannosidase alpha class 2A member 1
13	RPTN	Repetin
13	TCHHL1	trichohyalin like 1
13	S100A11	protein S100-A11
13	S100A10	Protein S100-A10
13	THEM4	thioesterase superfamily member 4
13	THEM5	thioesterase superfamily member 5
13	RORC	RAR related orphan receptor C
13	LINGO4	leucine rich repeat and Ig domain containing 4
13	TDRKH	tudor and KH domain containing
13	OAZ3	Oryctolagus cuniculus ornithine decarboxylase antizyme 3 (OAZ3), mRNA.
13	MRPL9	mitochondrial ribosomal protein L9
13	RIIAD1	regulatory subunit of type II PKA R-subunit (RIIa) domain containing 1
13	TNRC4	Trinucleotide Repeat-Containing Gene 4 Protein
13	SNX27	sorting nexin family member 27
13	TUFT1	tuftelin 1
13	CGN	cingulin
13	POGZ	pogo transposable element with ZNF domain
13	PSMB4	proteasome subunit beta type-4
13	SELENBP1	selenium binding protein 1

13	PI4KB	phosphatidylinositol 4-kinase beta
13	ZNF687	zinc finger protein 687
13	PSMD4	proteasome 26S subunit, non-ATPase 4
13	PIP5K1A	phosphatidylinositol-4-phosphate 5-kinase type 1 alpha
13	VPS72	vacuolar protein sorting-associated protein 72 homolog
13	TMOD4	tropomodulin 4
13	SCNM1	sodium channel modifier 1
13	LYSMD1	LysM domain containing 1
13	TNFAIP8L2	Tumor necrosis factor alpha-induced protein 8-like protein 2
13	SEMA6C	semaphorin-6C precursor
13	GABPB2	GA binding protein transcription factor beta subunit 2
13	MLLT11	myeloid/lymphoid or mixed-lineage leukemia; translocated to, 11
13	CDC42SE1	CDC42 small effector 1
13	C1orf56	chromosome 1 open reading frame 56
13	BNIP1	BCL2/adenovirus E1B 19kD interacting protein like
13	PRUNE1	prune exopolyphosphatase
13	FAM63A	family with sequence similarity 63 member A
13	ANXA9	annexin A9
13	CERS2	ceramide synthase 2
13	SETDB1	SET domain bifurcated 1
13	ARNT	aryl hydrocarbon receptor nuclear translocator
13	CTSK	Cathepsin K
13	CTSS	cathepsin S
13	ENSA	endosulfine alpha
13	MCL1	myeloid cell leukemia 1
13	ADAMTSL4	ADAMTS like 4
13	ECM1	extracellular matrix protein 1
13	RPRD2	regulation of nuclear pre-mRNA domain containing 2
13	PRPF3	pre-mRNA processing factor 3
13	CIART	circadian associated repressor of transcription
13	C1orf54	chromosome 1 open reading frame 54
13	APH1A	aph-1 homolog A, gamma-secretase subunit
13	CA14	carbonic anhydrase 14
13	PLEKHO1	pleckstrin homology domain containing O1
14	MBNL1	muscleblind like splicing regulator 1
17	PNRC2	proline-rich nuclear receptor coactivator 2
17	BMP4	bone morphogenetic protein 4
17	CDKN3	cyclin dependent kinase inhibitor 3
17	GMFB	glia maturation factor beta

17	CGRRF1	cell growth regulator with ring finger domain 1
17	SAMD4A	sterile alpha motif domain containing 4A
17	PTGDR	prostaglandin D2 receptor (DP)
17	PTGER2	prostaglandin E receptor 2
17	TXNDC16	thioredoxin domain containing 16
17	GPR137C	G protein-coupled receptor 137C
17	ERO1A	endoplasmic reticulum oxidoreductase alpha
17	PSMC6	proteasome 26S subunit, ATPase 6]
17	STYX	serine/threonine/tyrosine interacting protein
17	FERMT2	fermitin family member 2
17	DDHD1	DDHD domain containing 1
21	PES1	pescadillo ribosomal biogenesis factor 1
21	TCN2	transcobalamin 2
21	SLC35E4	solute carrier family 35 member E4
21	C5orf52	chromosome 5 open reading frame 52
21	OSBP2	oxysterol binding protein 2
21	MORC2	MORC family CW-type zinc finger 2
21	SMTN	smoothelin
21	INPP5J	inositol polyphosphate-5-phosphatase J
21	PLA2G3	phospholipase A2 group III
21	RNF185	ring finger protein 185
21	LIMK2	LIM domain kinase 2
21	PIK3IP1	phosphoinositide-3-kinase interacting protein 1
21	PATZ1	POZ/BTB and AT hook containing zinc finger 1
21	DRG1	developmentally regulated GTP binding protein 1
21	EIF4ENIF1	eukaryotic translation initiation factor 4E nuclear import factor 1
21	SFI1	SFI1 centrin binding protein
X	DACH2	dachshund family transcription factor 2

## DISCUSIÓN GENERAL Y CONCLUSIONES

El tamaño de camada mantiene el interés como carácter objetivo de selección en las diferentes especies prolíficas por su importancia económica. No obstante, es un carácter complejo por ser el resultado de un proceso secuencial desde la ovulación hasta el parto. La respuesta genética ha sido baja y el desconocimiento de su naturaleza genética hace difícil encontrar una estrategia de selección más óptima (Antonini et al. 2009; Noguera et al. 2002). El objetivo de nuestro estudio fue contribuir al conocimiento de las regiones genómicas asociadas al tamaño camada para generar información que pueda ser útil en la mejora genética de este carácter.

La detección de áreas genómicas asociadas al tamaño de camada y sus componentes se abordó mediante un estudio de asociación del genoma completo (GWA) y mediante la estimación de índices de diferenciación genética (Fst). El GWAS sólo puede detectar las áreas genómicas cuyos genes tuvieron un efecto mayoritario en los animales del experimento (Ball, 2013). En cambio, los índices de diferenciación genética detectan las áreas que con mayores frecuencias han sido afectadas por el proceso de las fuerzas evolutivas (selección y deriva en nuestro experimento) (Hedrick, 2011). Las áreas genómicas detectadas con ambas aproximaciones fueron distintas. El hecho de que no coincidan las áreas genómicas detectadas por GWAS y Fst, apoyaría la hipótesis de que la variabilidad genética observada para el tamaño de camada y sus componentes tiene una base poligénica grande. Con estas aproximaciones no se detectarían los genes de gran efecto, posiblemente ya fijados por la selección natural o artificial en el conejo doméstico (Carneiro et al., 2014b). En nuestro experimento de selección divergente se utilizaron tres poblaciones que provienen de una línea seleccionada previamente por tamaño de camada al destete.

Asumiendo que la variabilidad genética estuviera explicada por muchos genes, se deberían encontrar un número elevado de áreas genómicas y de genes asociados a los caracteres, sin embargo en nuestras líneas solo se encontraron 8 regiones genómicas que contienen 37 genes diferentes. Nuestro experimento tiene la limitación de que el número de animales utilizados es bajo, lo que dificulta la detección de un número grande de genes poligénicos. En porcino, utilizando un número mayor de animales que en nuestro estudio, se han detectado hasta 80 regiones genómicas asociadas al tamaño de camada y la tasa de ovulación ( $\approx$  400 animales, Onteru et al. 2011;  $\approx$  1000 animales, Schneider et al. 2014).

El análisis de asociación del genoma completo, utilizando datos del segundo parto, detectó regiones genómicas asociadas en los cromosomas 11, 14, 17 y 21. Especial atención tiene las regiones del cromosoma 17, ya que se solapan para el tamaño de camada, el número de embriones implantados y el número de nacidos vivos. Esto podría ser el resultado de la elevada correlación entre estos tres caracteres (García y Baselga, 2002b; Laborda et al., 2012b). Los genes candidatos dentro de las regiones genómicas asociadas al tamaño de camada pueden llegar a explicar el 21,84% de su varianza genética.

En el total de áreas genómicas detectadas por el GWAS se han identificado 37 genes, de los cuales presumimos que por lo menos la mitad de ellos son asociaciones falsas. Esto es respaldado por los resultados para la potencia de detección de asociaciones obtenidos en los estudios de simulación con las líneas divergentes. El estudio muestra que el valor predictivo positivo para 10,010 QTN fue del 57% por término medio. Esto quiere decir que aproximadamente la mitad de los SNPs dentro de las ventanas serían falsos positivos y no estarían asociados a los caracteres estudiados. El modelo infinitesimal (10,010 QTN) supuesto en el estudio de simulación mostró que a pesar de tener un moderado a alto error tipo I (falsos positivos), tendríamos una potencia de detección y un valor predictivo positivo aceptables. Esta alta potencia utilizando líneas divergentes se debe a que los QTNs se encuentran con un alto valor para la menor frecuencia alélica (MAF). Un MAF elevado aumenta la potencia de detección de asociación de un QTN sin importar la magnitud del efecto de QTN (De Maturana et al., 2014)

Las huellas de selección detectadas podrían ser debido al efecto de la selección y/o la deriva génica (Grams et al., 2015). La deriva puede ser un factor importante en los procesos de selección dependiendo del tamaño efectivo de la población. Los resultados de la simulación con el mismo tamaño efectivo de la población del experimento de selección por capacidad uterina demostraron que por deriva se fijan hasta un 50% de los QTNs. El efecto de la deriva se estimó que podría suponer hasta el 14.52% de la respuesta genética a la selección por capacidad uterina.

En síntesis, la investigación tiene como conclusiones generales, los siguientes enunciados:

A. Estudio de simulación del experimento de selección divergente por capacidad uterina

- Con el mismo número de individuos, la estrategia de utilizar animales procedentes de dos líneas divergentes por capacidad uterina es más eficaz para detectar asociaciones positivas, entre regiones genómicas y el carácter seleccionado, que cuando se utilizan animales de una sola línea.
- Los tres escenarios de estructura genética (10010, 1010, 110 QTNs) analizados con las líneas divergentes presentan valores similares para el error tipo I y la potencia de detección. Sin embargo, el grupo de 10,010 QTNs presenta un valor por término medio más elevado para el valor predictivo positivo.
- La deriva tiene un efecto moderado, entre un 35 a un 50% de los QTNs fijados en el experimento de selección divergente por capacidad uterina. Estos QTNs fijados solo explican entre 10% y el 15% de la respuesta de selección por capacidad uterina.

B. Análisis por GWAS y huellas de selección del experimento de selección divergente:

- Se detectaron 4, 4, 1 y 5 regiones genómicas asociadas al tamaño de camada, número de nacidos vivos, tasa de ovulación y número de embriones implantados, respectivamente. Para tamaño de camada, las cuatro ventanas relevantes explican el 21.84 % de la varianza genética y el 3.55% de la varianza fenotípica. El 64% de la varianza genética explicada por estas ventanas se debe a los QTL del cromosoma 17.
- Se identificaron una lista de 37 genes dentro de las ventanas asociadas al tamaño de camada, el número de nacidos vivos, la tasa de ovulación y el número de embriones implantados. Los genes BMP4, LIMK2, PTGDR, STYX, ERO1A, ANKH y CDKN3 se perfilan como posibles genes candidatos para el tamaño de camada y sus componentes.
- Se encontraron como huellas de selección, tres “clúster” relevantes entre la línea de alta capacidad uterina y la línea V, y seis “clúster” entre la línea de baja capacidad uterina y la línea V. No se encontraron “clúster” entre las líneas divergentes con los criterios establecidos en este estudio.

## REFERENCIAS BIBLIOGRÁFICAS

- Anderson, C. a, Pettersson, F.F.H., Clarke, G.G.M., Cardon, L.R., Morris, A.P., y Zondervan, K.T. (2010). Data quality control in genetic case-control association studies. *Nature protocols* 5, 1564–1573.
- Antonini, A.G., Corva, S.G., y Oyarzabal, M.I. (2009). Selección por tamaño de camada en conejos para carne. *Journal of Basic & Applied Genetics* 19, 1–7.
- Argente, M.J. (1996). Selección Divergente por Eficiencia Uterina en Conejo. Universidad Politecnica de Valencia.
- Argente, M.J., Santacreu, M.A., Climent, A., Bolet, G., y Blasco, A. (1997). Divergent selection for uterine capacity in rabbits. *Journal of Animal Science* 75, 2350–2354.
- Argente, M.J., Santacreu, M.A., Climent, A., y Blasco, A. (2003a). Relationships between uterine and fetal traits in rabbits selected on uterine capacity. *Journal of Animal Science* 81, 1265–1273.
- Argente, M.J., Blasco, A., Ortega, J.A., Haley, C.S., y Visscher, P.M. (2003b). Analyses for the presence of a major gene affecting uterine capacity in unilaterally ovariectomized rabbits. *Genetics* 163, 1061–1068.
- Argente, M.J., Merchán, M., Peiró, R., García, M.L., Santacreu, M. a., Folch, J.M., y Blasco, A. (2010). Candidate gene analysis for reproductive traits in two lines of rabbits divergently selected for uterine capacity. *Journal of Animal Science* 88, 828–836.
- Badawy, A.Y., Peiró, M., y Santacreu, M.A. (2013). Efecto de la Superovulación Sobre la Supervivencia Embrionaria y Fetal en Conejas Multíparas. In XV JJornadas sobre Producción Animal, (Zaragoza, España), p.
- Baena, P.L., García, M.L., Muelas, R., Agea, I., Rodríguez, B., y Argente, M.J. (2006). Efecto del Estado Fisiológico y la Estación sobre los Caracteres Reproductivos Durante el Periodo de Lactación en Conejo. XXXI Symposium de Cunicultura 139–144.
- Ball, R.D. (2013). Designing a GWAS: Power, sample size, and data structure. *Methods in Molecular Biology* 1019, 37–98.
- Ballester, M., Castelló, A., Peiró, R., Argente, M.J., Santacreu, M.A., y Folch, J.M. (2013). Identification of differentially expressed genes in the oviduct of two rabbit lines divergently selected for uterine capacity using suppression subtractive hybridization. *Animal genetics* 44, 296–304.
- Baselga, M. (2001). Selección del tamaño de camada en conejo: posibilidades de seleccion por hiperprolificidad. In XXV Symposium de Cunicultura, (Asociación Española de

- Cunicultura (ASESCU)), pp. 69–66.
- Baselga, M., y Blasco, A. (1989). *Mejora Génética del Conejo de Producción de Carne* (Madrid, España: Ediciones Mundi-Prensa).
- Bazer, F.W., Robinson, O.W., Clawson, A.J., y Ulberg, L.C. (1969). Uterine Capacity at Two Stages of Gestation in Gilts Following Embryo Superinduction. *Journal of Animal Science* 29, 30–34.
- Bennett, G.L., y Leymaster, K.A. (1990). Genetic Implications of a simulation Model of Litter Size in Swine Based on Ovulation Rate, Potential Embryonic Viability and Uterine Capacity: II Simulated Selection. *Journal Animal Science* 68, 980–986.
- Bidanel, J., Gruand, J., y Legault, C. (1996a). Genetic variability of age and weight at puberty, ovulation rate and embryo survival in gilts and relations with production traits. *Genetics, Selection, Evolution : GSE* 28, 103–115.
- Bidanel, J.P., Blasco, A., Gogué, J., Lagant, H., y Avord, D.D.G. (1996b). Résultats de quatre générations de sélection pour le taux d'ovulation et la survie prénatale chez des porcs de race Large White. *Journées de la Recherche Porcine en France* 28, 1–8.
- Blasco, A., y Toro, M.A. (2014). A short critical history of the application of genomics to animal breeding. *Livestock Science* 166, 4–9.
- Blasco, A., Bidanel, J.P., Bolet, G., Haley, C.S., y Santacreu, M.A. (1993). The genetics of prenatal survival of pigs and rabbits: a review. *Livestock Production Science* 37, 1–21.
- Blasco, A., Argente, M.J., Haley, C.S., y Santacreu, M.A. (1994). Relationships Between Components of Litter Size in Unilaterally Ovariectomized and Intact Rabbit Does. *Journal of Animal Science* 72, 3066–3072.
- Blasco, A., Bidanel, J.P., y Haley, C.S. (1995). Genetics and Neonatal Survival. In *The Neonatal Pig: Development and Survival*, CAB International (Royaume Uni), ed. (Wallingford: CAB Int), pp. 17–35.
- Blasco, A., Ortega, J.A., Climent, A., y Santacreu, M.A. (2005). Divergent Selection for Uterine Capacity in Rabbits . I . Genetic Parameters and Response to Selection. *Journal Animal Science* 83, 2297–2302.
- Bolet, G., y Theau-Clément, M. (1994). Fertilisation rate and preimplantation embryonic development in two rabbit strains of different fecundity, in purebreeding and crossbreeding. *Animal Reproduction Science* 36, 153–162.
- Bolet, G., Santacreu, M.A., Argente, M.J., Climent, A., y Blasco, A. (1994). Divergent Selection for Uterine Efficiency in Unilaterally Ovariectomized Rabbits. I. Phenotypic and Genetic Parameters. In *5th World Congress on genetics applied to livestock*

- production., (Guelph), pp. 261–264.
- Bradford, G.E. (1969). Genetic control of ovulation rate and embryo survival in mice. I. Response to selection. *Genetics* 61, 905–921.
- Bradford, G.E. (1979). Genetic Variation in Prenatal Survival and Litter Size. *Journal Animal Science* 49, 66–74.
- Bush, W.S., y Moore, J.H. (2012). Chapter 11: Genome-wide association studies. *PLoS computational biology* 8, e1002822.
- Carneiro, M., Afonso, S., Geraldés, A., Garreau, H., Bolet, G., Boucher, S., Tircazes, A., Queney, G., Nachman, M.W., y Ferrand, N. (2011). The Genetic Structure of Domestic Rabbits. *Molecular Biology and Evolution* 28, 1801–1816.
- Carneiro, M., Albert, F.W., Afonso, S., Pereira, R.J., Burbano, H., Campos, R., Melo-Ferreira, J., Blanco-Aguiar, J.A., Villafuerte, R., Nachman, M.W., et al. (2014a). The Genomic Architecture of Population Divergence between Subspecies of the European Rabbit. *PLoS Genetics* 10, e1003519.
- Carneiro, M., Rubin, C., Di Palma, F., Albert, F.W., Alfoldi, J., Barrio, A.M., Pielberg, G., Rafati, N., Sayyab, S., Turner-Maier, J., et al. (2014b). Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* 345, 1074–1079.
- Cartuche, L., Pascual, M., Gómez, E.A., y Blasco, A. (2014). Economic weights in rabbit meat production. *World Rabbit Science* 22, 165–177.
- Castellini, C., Dal Bosco, A., Arias-Álvarez, M., Lorenzo, P.L., Cardinali, R., y Rebollar, P.G. (2010). The main factors affecting the reproductive performance of rabbit does: A review. *Animal Reproduction Science* 122, 174–182.
- Christenson, R.K., Leymaster, K.A., y Young, L.D. (1987). Justification of unilateral hysterectomy-ovariectomy as a model to evaluate uterine capacity in swine. *Journal of animal science* 65, 738–744.
- Clutter, A.C., Nielsen, M.K., y Johnson, R.K. (1990). Alternative of selection for Litter Size in mice: I. Characterization of Base Population and Development of Methods. *Journal of Animal Science* 68, 3536–3542.
- Clutter, A.C., Kochera Kirby, Y.L., y Nielsen, M.K. (1994). Uterine capacity and ovulation rate in mice selected 21 generations on alternative criteria to increase litter size. *Journal of animal science* 72, 577–583.
- Cockett, N.E. (2010). Reproductive Genomics: Genome, Transcriptome, and Proteome Resources. In *Reproductive Genomics in Domestic Animals*, (Oxford, UK: Wiley-Blackwell), pp. 3–22.

- Demars, J., Fabre, S., Sarry, J., Rossetti, R., Gilbert, H., Persani, L., Tosser-Klopp, G., Mulsant, P., Nowak, Z., Drobik, W., et al. (2013). Genome-Wide Association Studies Identify Two Novel BMP15 Mutations Responsible for an Atypical Hyperproliferacy Phenotype in Sheep. *PLoS Genetics* 9, e1003482.
- Du, F.-X., Clutter, A.C., y Lohuis, M.M. (2007). Characterizing linkage disequilibrium in pig populations. *International journal of biological sciences* 3, 166–178.
- Egena, S.S.A., Akpa, G.N., Alemede, I. C., y Aremu, A. (2012). Genetic and Non-Genetic Factors Affecting Litter Size and Birth Weight of Rabbit in Minna , Niger State , Nigeria. *Animal Production* 14, 160–166.
- Elmaghraby, M.M.A., y Elkholya, S.Z. (2010). Characterizing Litter Losses in Purebred New Zealand White Rabbits. *Lucr ri tiin ifice-Universitatea de tiin e Agricole i Medicin Veterinar* 54, 304–310.
- Falconer, D.S., y Mackay, T.F.C. (1996). *Introducción a la Genética Cuantitativa* (Zaragoza, España: Editorial ACRIBIA).
- Fayeye, T.R., y Ayorinde, K.L. (2010). Effects of Season , Generation , Number of Mating , Parity and Doe Number of Teat on Doe and Litter Birth Characteristics in Domestic Rabbit. *Research Journal of Animal and Veterinary Sciences* 5, 6–9.
- Fernando, R.L., y Garrick, D. (2013). Bayesian methods applied to GWAS. In *Genome-Wide Association Studies and Genomic Prediction*, C. Gondro, J.H. van der Werf, y B. Hayes, eds. (Human Press Inc.), pp. 237–274.
- Ferraz, J., y Eler, J. (1994). Use of different animal models in prediction of genetic parameters of 23 traits of Californian and New Zealand White rabbits raised in tropics and suggestion of Selection Criteria. In *5th World Congress on genetics applied to livestock production.*, (Guelph, Canada), pp. 348–351.
- Fortun, L., Prunier, A., Étienne, M., y Lebas, F. (1994). Influence of the nutritional deficit of foetal survival and growth and plasma metabolites in rabbit does. *Reproduction Nutrition Development* 34, 201–211.
- Foxcroft, G.R., y Hasnain, H. (1973). Embryonic Mortality in the Post-Parturient Domestic Rabbit. *Journal of Reproduction and Fertility* 33, 315–318.
- Freking, B.A., Leymaster, K.A., Vallet, J.L., y Christenson, R.K. (2007). Number of fetuses and conceptus growth throughout gestation in lines of pigs selected for ovulation rate or uterine capacity. *Journal of Animal Science* 85, 2093–2103.
- Freking, B.A., Lents, C.A., y Vallet, J.L. (2016). Selection for uterine capacity improves lifetime productivity of sows. *Animal Reproduction Science* 167, 16–21.
- Garcia, F., y Perez, A. (1989). Efectos de la lactacion y numero de lactantes sobre la monta,

- ovulación y supervivencia fetal hasta el parto, evaluados por laparoscopia, en conejas multiparas. *Inf. Tec. Econ. Agraria*, 80: 3-3, 10.
- García, F., Baselga, M., Pla, M., y Deltoro, J. (1984). Ovulation Rate in Meat Rabbit. 3rd World Rabbit Congress 2, 436.
- García, M.L., y Baselga, M. (2002a). Estimation of Genetic Response to selection in litter size of rabbits using a cryopreserved control population. *Livestock Production Science* 74, 45–53.
- García, M.L., y Baselga, M. (2002b). Genetic Response to Selection for Reproductive Performance in a Maternal Line of Rabbits. *World Rabbit Science* 10, 71–76.
- García, M.L., Peiró, R., Argente, M.J., Merchán, M., Folch, J.M., Blasco, A., y Santacreu, M.A. (2010). Investigation of the oviductal glycoprotein 1 (OVGP1) gene associated with embryo survival and development in the rabbit. *Journal of Animal Science* 88, 1597–1602.
- García-Ximénez, F., y Vicente, J.S. (1992). Effect of ovarian cystic haemorrhagic follicles embryo recovery and survival after transfer in hCG-ovulated rabbits. *Reproduction Nutrition Development* 32, 143–149.
- Gion, J.M., Clutter, A.C., y Nielsen, M.K. (1990). Alternative methods of selection for litter size in mice. II. Response to 13 generations of selection. *J. Anim. Sci.* 68, 3543–3556.
- Grams, V., Wellmann, R., Preuß, S., Grashorn, M.A., Kjaer, J.B., Bessei, W., y Bennewitz, J. (2015). Genetic parameters and signatures of selection in two divergent laying hen lines selected for feather pecking behaviour. *Genetics Selection Evolution* 47, 77.
- Hayes, B. (2013). Overview of statistical methods for genome-wide association studies (GWAS). In *Genome-Wide Association Studies and Genomic Prediction*, C. Gondro, J.H. van der Werf, y B. Hayes, eds. (Humana Press), pp. 149–169.
- Hedrick, P.W. (2011). *Genetics of Population* (Jones and Barlet Publishers, LLC).
- Hernández, S.C. (2011). Genetics of litter size and prenatal survival in pigs.
- Hickey, J.M., y Gorjanc, G. (2013). Simulated Data for Genomic Selection and Genome-Wide Association Studies Using a Combination of Coalescent and Gene Drop Methods. *G3 Genes | Genomes | Genetics* 2, 425–427.
- Iraqi, M.M., Afifi, E.A., Nayera, Z.B., y Gad, S.M.A. (2007). Estimation of genetic parameters for litter traits in gabali rabbits raised in the north-western coast of egypt using multi-trait animal model. In *5th International Conference on Rabbit Production*, (Hurghada, Egypt), pp. 103–112.
- Johansson, A.M., Pettersson, M.E., Siegel, P.B., y Carlborg, Ö. (2010). Genome-Wide Effects

- of Long-Term Divergent Selection. *PLoS Genetics* 6, e1001188.
- Johnson, R.K., Nielsen, M.K., y Casey, D.S. (1999). Responses in Ovulation Rate, Embryonal Survival, and Litter Traits in Swine to 14 Generations of Selection to Increase Litter Size. *Journal of Animal Science* 77, 541–557.
- Khalil, M.H. (1993). Diversity of Repeatability between Parities for Litter Traits and Reproductive Interval in Doe Rabbits. *World Rabbit Science* 1, 147–154.
- Khalil, M.H., y Al-Saef, A.M. (2008). Methods, criteria, techniques and genetic responses for rabbit selection: A review. *Proc. 9th World Rabbit Congress* 1–22.
- Khalil, M.H., Owen, J.B., y Afifi, E.A. (1987). A genetic analysis of litter traits in Bauscat and Giza White rabbits. *Animal Production* 45, 123–134.
- Kimura, M., y Ohta, T. (1971). *Theoretical aspects of population genetics* (Princeton University Press).
- Kirby, Y.K., y Nielsen, M.K. (1993). Alternative Methods of Selection for Litter Size in Mice: III. Response to 21 Generations of Selection. *Journal Animal Science* 71, 571–578.
- Laborda, P., Mocé, M.L., Santacreu, M.A., y Blasco, A. (2011). Selection for ovulation rate in rabbits: Genetic parameters, direct response, and correlated response on litter size. *Journal of Animal Science* 89, 2981–2987.
- Laborda, P., Santacreu, M.A., Blasco, A., y Moce, M.L. (2012a). Selection for ovulation rate in rabbits: Direct and correlated responses estimated with a cryopreserved control population. *Journal of Animal Science* 90, 3392–3397.
- Laborda, P., Mocé, M.L., Blasco, A., y Santacreu, M.A. (2012b). Selection for ovulation rate in rabbits: Genetic parameters and correlated responses on survival rates. *Journal of Animal Science* 90, 439–446.
- Lamberson, W.R., Johnson, R.K., Zimmerman, D.R., y Long, T.E. (1991). Direct responses to selection for increased litter size, decreased age at puberty, or random selection following selection for ovulation rate in swine. *Journal of animal science* 69, 3129–3143.
- Lambert, E., Williams, D.H., Lynch, P.B., Hanrahan, T.J., McGeady, T.A., Austin, F.H., Boland, M.P., y Roche, J.F. (1991). The extent and timing of prenatal loss in gilts. *Theriogenology* 36, 655–665.
- Ledur, M.C., Navarro, N., y Pérez-Enciso, M. (2010). Large-scale SNP genotyping in crosses between outbred lines: how useful is it? *Heredity* 105, 173–182.
- Leymaster, K.A., y Christenson, R. (2000). Direct and correlated responses to selection for ovulation rate or uterine capacity in swine. *Journal Animal Science* 8, 68 (abstract # 284).

- Maertens, L. (1998). Effect of Flushing, Mother-Litter Separation and Pmsg on the Fertility of Lactating Does and the Performance of Their Litter. *World Rabbit Science* 6, 185–190.
- Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. *Nature* 461, 747–753.
- Martínez-Álvaro, M. (2013). Divergent selection for residual variance of litter size in rabbits. Universidad Politécnica de Valencia.
- De Maturana, E.L., Ibáñez-Escriche, N., González-Recio, Ó., Marenne, G., Mehrban, H., Chanock, S.J., Goddard, M.E., y Malats, N. (2014). Next generation modeling in GWAS: comparing different genetic architectures. *Human Genetics* 133, 1235–1253.
- McNitt, J.I., Lukefahr, S.D., Cheeke, P.R., y Patton, N.M. (2013). *Rabbit Production* (Boston: CABI Wallingford and Boston).
- Merchán, M., Peiró, R., Estellé, J., Sastre, Y., Santacreu, M.A., y Folch, J.M. (2004). Análisis de los genes candidatos de la oviductina, uteroglobina y receptor de progesterona en dos líneas divergentes seleccionadas por capacidad uterina en conejo. *ITEA. Producción Animal* 100A, 234–237.
- Merchán, M., Peiró, R., Argente, M.J., Santacreu, M.A., García, M.L., Blasco, A., y Folch, J.M. (2009). Analysis of the oviductal glycoprotein 1 polymorphisms and their effects on components of litter size in rabbits. *Animal Genetics* 40, 756–758.
- Mesa, H., Safranski, T.J., Fischer, K. a, Cammack, K.M., y Lamberson, W.R. (2005). Selection for placental efficiency in swine: genetic parameters and trends. *J Anim Sci* 83, 983–991.
- Mocé, M.L., Santacreu, M.A., Climent, A., y Blasco, A. (2004). The effect of divergent selection for uterine capacity on fetal and placental development at term in rabbits: Maternal and embryonic genetic effects. *Journal of Animal Science* 82, 1046–1052.
- Mocé, M.L., Santacreu, M.A., Climent, A., y Blasco, A. (2005). Divergent selection for uterine capacity in rabbits . III . Responses in uterine capacity and its components estimated with a cryopreserved control population. *Journal of Animal Science* 83, 2308–2312.
- Moura, A.S.A.M., Costa, A.R., y Polastre, R. (2001). Variance Components and Response to Selection for Reproductive, Litter and Growth Traits Through a Multi-Purpose Index. *World Rabbit Science* 9, 77–86.
- Nagy, I., Farkas, J., Curik, I., Gorjanc, G., Gyovai, P., y Szendrő, Z. (2014). Estimation of

- additive and dominance variance for litter size components in rabbits. *Czech Journal of Animal Science* 59, 182–189.
- Nielsen, M.K., Kittok, R.J., y Kochera Kirby, Y.L. (1995). Uterine mass and uterine blood volume in mice selected 21 generations for alternative criteria to increase litter size. *Journal of animal science* 73, 2243–2248.
- Noguera, J.L., Varona, L., Babot, D., y Estany, J. (2002). Multivariate analysis of litter size for multiple parities with production traits in pigs: Bayesian Variance Component Estimation. *Journal of Animal Science* 80, 2548–2555.
- Oliver, M.A., Gispert, M., y Diestre, A. (1993). The effects of breed and halothane sensitivity on pig meat quality. *Meat Science* 35, 105–118.
- Onteru, S.K., Fan, B., Du, Z.-Q., Garrick, D.J., Stalder, K.J., y Rothschild, M.F. (2011a). A whole-genome association study for pig reproductive traits. *Animal Genetics* 43, 18–26.
- Onteru, S.K., Fan, B., Nikkilä, M.T., Garrick, D.J., Stalder, K.J., y Rothschild, M.F. (2011b). Whole-genome association analyses for lifetime reproductive traits in the pig. *Journal of Animal Science* 89, 988–995.
- Peiro, R., Herrler, A., Santacreu, M.A., Merchan, M., Argente, M.J., Garcia, M.L., Folch, J.M., y Blasco, A. (2010). Expression of progesterone receptor related to the polymorphism in the PGR gene in the rabbit reproductive tract. *Journal of Animal Science* 88, 421–427.
- Peiró, R., Merchán, M., Santacreu, M.A., Argente, M.J., García, M.L., Folch, J.M., y Blasco, A. (2008). Identification of single-nucleotide polymorphism in the progesterone receptor gene and its association with reproductive traits in rabbits. *Genetics* 180, 1699–1705.
- Piles, M., García, M.L., Rafel, O., Ramon, J., y Baselga, M. (2006). Genetics of litter size in three maternal lines of rabbits: Repeatability versus multiple-trait models. *Journal of Animal Science* 84, 2309–2315.
- Poujardieu, B., y Théau-Clément, M. (1995). Productivité de la lapine et état physiologique. *Annales de Zootechnie, INRA/EDP Sciences* 44, 29–39.
- Pryce, J.E., y Daetwyler, H.D. (2012). Designing dairy cattle breeding schemes under genomic selection: a review of international research. *Animal Production Science* 52, 107–114.
- Ragab, M., Sánchez, J.P., Mínguez, C., Vicente, J.S., y Baselga, M. (2014). Litter size components in a full diallel cross of four maternal lines of rabbits. *Journal Animal Science* 92, 3231–3236.

- Ribeiro, E.L.D.A., Nielsen, M.K., Bennett, G.L., y Leymaster, K.A. (1997). A Simulation Model Including Ovulation Rate, Potential Embryonic Viability, and Uterine Capacity to Explain Litter Size in Mice: I. Model Development and Implementation. *Journal of Animal Science* 75, 641–651.
- Rosendo, A., Druet, T., Gogu e, J., y Bidanel, J.P. (2007). Direct responses to six generations of selection for ovulation rate or prenatal survival in Large White pigs. *Journal of Animal Science* 85, 356–364.
- Rothschild, M., Jacobson, C., Vaske, D., Tuggle, C., Wang, L., Short, T., Eckardt, G., Sasaki, S., Vincent, A., McLaren, D., et al. (1996). The estrogen receptor locus is associated with a major gene influencing litter size in pigs. *Proceedings of the National Academy of Sciences of the United States of America* 93, 201–205.
- Ru ız-Flores, A., y Johnson, R.K. (2001). Direct and correlated responses to two-stage selection for ovulation rate and number of fully formed pigs at birth in swine. *Journal of Animal Science* 79, 2286–2297.
- Santacreu, M.A. (1992). Estimaci n de los parametros gen ticos de la tasa de ovulaci n, supervivencia prenatal y tama o de camada en conejo. Tesis. Universidad Politecnica de Valencia.
- Santacreu, M.A. (2006). La Supervivencia Prenatal en la Coneja Reproductora. XXXI Symposium de Cunicultura 243–250.
- Santacreu, M., Viudes de Castro, M.P., y Blasco, A. (1990). Evaluation par coelioscopie des corps jaunes et des embryons. Influence sur la taille de port e chez la lapine. *Reproduction Nutrition Development* 30, 583–588.
- Santacreu, M.A., Climent, A., Gallego, M., Fayos, L., y Blasco, A. (1996). Fertilization rate and early embryo development in two rabbit lines selected on uterine capacity. In 6th World Rabbit Congress, (Toulouse, France), pp. 355–357.
- Santacreu, M.A., Moc e, M.L., Climent, A., y Blasco, A. (2005). Divergent selection for uterine capacity in rabbits . II. Correlated Response in Litter Size and its components estimated with a cryopreserved control population. *Journal of Animal Science* 83, 2303–2307.
- Schneider, J.F., Rempel, L. a., Snelling, W.M., Wiedmann, R.T., Nonneman, D.J., y Rohrer, G.A. (2012). Genome-wide association study of swine farrowing traits. Part II: Bayesian analysis of marker data,. *Journal of Animal Science* 90, 3360–3367.
- Schneider, J.F., Nonneman, D.J., Wiedmann, R.T., Vallet, J.L., y Rohrer, G.A. (2014). Genomewide association and identification of candidate genes for ovulation rate in swine. *Journal of Animal Science* 92, 3792–3803.

- Schneider, J.F., Miles, J.R., Brown-Brandl, T.M., Nienaber, J.A., Rohrer, G.A., y Vallet, J.L. (2015). Genomewide association analysis for average birth interval and stillbirth in swine. *Journal of Animal Science* 93, 529.
- Smidt, V.D., Thomsen, J., y Hafez E.S.E. (1968). Investigation on the influence of lactation on implantation and foetal development in the rabbit. In *Zeitschrift fuer tierzuechtung und zuechtungsbiologie (Journal of animal breeding and genetics), (Z. Tierzücht Zü chtbiol)*, pp. 115–126.
- Sorensen, P., Kjaer, J.B., Brenoe, U.T., y Su, G. (2001). Estimates of Genetic Parameters in Danish White Rabbits Using an Animal Model: II. Litter Traits. *World Rabbit Science* 9, 33–38.
- Spötter, A., y Distl, O. (2006). Genetic approaches to the improvement of fertility traits in the pig. *The Veterinary Journal* 172, 234–247.
- Sternstein, I., Reissmann, M., Maj, D., Bieniek, J., y Brockmann, G.A. (2015). A comprehensive linkage map and QTL map for carcass traits in a cross between Giant Grey and New Zealand White rabbits. *BMC Genetics* 16, 16.
- Stram, D.O. (2014). *Design, Analysis, and Interpretation of Genome-Wide Association Scans* (New York, NY: Springer New York).
- Torrès, S., Meunier, M., y Sevellec, C. (1987). Comparative study of preimplantation development embryonic loss in two rabbit strains. *Reproduction Nutrition Development* 27, 707–714.
- Turner, S., Armstrong, L.L., Bradford, Y., Carlson, C.S., Dana, C., Crenshaw, A.T., Andrade, M. De, Doheny, K.F., Jonathan, L., Hayes, G., et al. (2011). Quality control procedures for genome wide association studies. *Current Proceedings in Human Genetics* 68, 1–24.
- Vallet, J.L., Klemcke, H.G., y Christenson, R.K. (2002). Interrelationships among conceptus size, uterine protein secretion, fetal erythropoiesis, and uterine capacity. *Journal of Animal Science* 80, 729–737.
- Vallet, J.L., Nonneman, D.J., y Kuehn, L.A. (2010). Quantitative Genomics of Female Reproduction. In *Reproductive Genomics in Domestic Animals*, (Oxford, UK: Wiley-Blackwell), pp. 23–51.
- Viudes De Castro, M.P., Cortell, C., Mocé, E., Marco-Jiménez, F., Joly, T., y Vicente, J.S. (2009). Effect of recombinant gonadotropins on embryo quality in superovulated rabbit does and immune response after repeated treatments. *Theriogenology* 72, 655–662.
- Wakeley, J. (2009). *Coalescent Theory: An Introduction* (Colorado).

Wilmot, I., Sales, D.I., y Ashworth, C.J. (1986). Maternal and embryonic factors associated with prenatal loss in mammals. *Reproduction* 76, 851–864.