



Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>



**Departamento de Ciencia Animal y de los
Alimentos**

Facultad de Veterinaria

Universidad Autónoma de Barcelona

**MODELIZACIÓN DE LA VARIABILIDAD MUTACIONAL SEGÚN LA
EDAD DE LOS PROGENITORES EN DOS ESPECIES DE MAMÍFEROS**

TESIS DOCTORAL

Mayela del Carmen Castillo Ojeda

Programa de Doctorado en Producción Animal

Bellaterra (Barcelona) 2017



**Departamento de Ciencia Animal y de los
Alimentos**

Facultad de Veterinaria

Universidad Autónoma de Barcelona

El trabajo de investigación “Modelización de la variabilidad mutacional según la edad de los progenitores en dos especies de mamíferos” que ha estado realizado por Mayela del Carmen Castillo Ojeda y dirigido por el Dr. Joaquim Casellas Vidal del Departamento de Ciencia Animal y de los Alimentos de la Universidad Autónoma de Barcelona, dentro del programa de doctorado de Producción Animal, se presenta como requisito para la obtención del grado de doctor.

Bellaterra, a septiembre de dos mil diecisiete.

El director de tesis

La doctoranda

Dr. Joaquim Casellas Vidal

Mayela del Carmen Castillo Ojeda

El pensamiento poblacional
Hay grandeza en esta concepción de la vida, ...
Que mientras este planeta ha ido girando
según la constante ley de la gravitación,
se han desarrollado y se están desarrollando,
a partir de un comienzo tan sencillo,
infinidad de formas cada vez más bellas y maravillosas.

Charles R. Darwin (1809-1882)

RESUMEN

La relevancia del impacto de las nuevas mutaciones sobre la variabilidad genética infinitesimal queda fuera de toda duda. No obstante, los mecanismos biológicos implicados en la generación de esta fuente de variabilidad genética están poco estudiados, incluso cuando algunos factores como la edad de los progenitores pueden ser de crucial importancia. Desafortunadamente, carecemos de métodos analíticos adecuados para abordar esta problemática, sobretodo cuando nos alejamos de los diseños experimentales típicos. Dentro de este contexto, la presente tesis doctoral se ha centrado precisamente en desarrollar una metodología de análisis específica, sobre la base de los modelos mixtos lineales. Esta metodología permite estimar la base genética de los caracteres cuantitativos diferenciando entre la variabilidad genética aditiva preexistente en la generación de individuos fundadores, y la nueva variabilidad genética originada por las mutaciones acaecidas en las distintas generaciones de la genealogía. Además, el componente mutacional se subdivide en función de su origen paterno o materno, permitiendo que la variabilidad genética mutacional se incremente (o reduzca) en función de la edad de cada progenitor. Para validar la utilidad de esta metodología analítica se han analizado dos bases de datos independientes, una centrada en el peso al destete de 12.644 ratones de la estirpe C57BL/6J y la otra considerando el peso al nacimiento de 8.130 terneros de la raza Bruna de los Pirineos. Con el objetivo de validar las distintas estructuras posibles del modelo de análisis se han implementado dos métodos estándar de comparación de modelos, el *deviance information criterion* y el factor de Bayes, ambos implementados dentro de un contexto de inferencia Bayesiana. Los análisis en ambas especies revelaron un comportamiento muy parecido en términos de varianzas genéticas mutacionales. Las dos poblaciones revelaron varianzas mutacionales tanto paternas como maternas pequeñas pero distintas de cero, claramente inferiores a la varianzas genéticas aditivas presentes en los fundadores. Aunque la varianzas mutacionales paternas se sugería ligeramente superior a la materna, los modelos de análisis no alcanzaron a detectar diferencias a nivel estadístico entre ambas fuentes de variación. Donde si se reportó un comportamiento claramente diferencial fue en la posibilidad de evolucionar con la edad de los progenitores. Tanto para los ratones C57BL/6J como para los terneros de raza Bruna de los Pirineos, la varianzas mutacionales paternas se incrementó de manera lineal con la edad del padre, llegando incluso a incrementos de casi el 50% durante el primer año de vida. Por el contrario, el componente mutacional materno no se vio afectado por la edad de la madre, manteniéndose constante durante toda su vida reproductiva. En conjunto, estos resultados deben verse como un avance significativo dentro del marco de investigación del fenómeno mutacional en mamíferos, con importantes implicaciones tanto para las especies ganaderas como para las de animales de laboratorio.

SUMMARY

The relevance of new mutations on the infinitesimal genetic variability remains out of any doubt. Nevertheless, biological mechanisms underlying this source of genetic variance are poorly understood, even when some factors such as parents' age may play a key role. Unfortunately, we lack of appropriate analytical approaches to investigate these phenomena, this shortage being specially dramatic when departing from standard experimental designs. This Ph.D. dissertation precisely focuses on the development of a statistical methodology for mutation analysis within the context of mixed linear models. This approach accounts for the genetic background of quantitative traits by making inference on two independent compounds, the additive genetic variability holding in the founder generation, and the new additive genetic variability originated by new mutations across the whole pedigree. Moreover, the mutational component differentiates between paternal and maternal origins, and may increase (or decrease) depending on parents' age. In order to validate this methodology, two different data sets were analyzed. The first accounted for weaning weight from 12,644 C57BL/6J mice, whereas the other one relied on the birth weight of 8,130 beef calves of the *Bruna dels Pirineus* breed. The different parameterizations were compared on the basis of two well known comparison criteria within the context of Bayesian inference, the deviance information criterion and the Bayes factor. The analyses revealed a very similar behavior in mice and calves. Both paternal and maternal variance component were small and different from zero. Although paternal mutational variance seemed slightly bigger than the maternal one, the analyses did not reveal statistical departures. Nevertheless, clearly different patterns were observed when considering parents' age. Both C57BL/6J mice and *Bruna dels Pirineus* calves evidenced a relevant gain of the parental mutation variance with sire's age, even reaching a 50% increase during the first age of life. On the contrary, the maternal mutational variance was not influenced by mother's age and kept stable along reproductive life. As a whole, these results must be viewed as a relevant advance within the research framework of mutations in mammals, they contributing remarkable implications for both laboratory and livestock species.

ÍNDICE

Introducción	1
1. Los ácidos nucleicos en el contexto evolutivo	3
1.1. Definición, tipos y características	3
1.2. Estructura y función del ADN	4
1.3. Replicación y transcripción	8
1.4. Traducción: Síntesis de proteínas	9
1.5. El ciclo celular	11
1.5.1. Mitosis	12
1.5.2. Meiosis	14
2. La variabilidad genética y las mutaciones	20
2.1. Variabilidad genética	20
2.2. Mutaciones	21
2.2.1. Clasificación de las mutaciones	23
2.3. Recurrencia de las mutaciones	26
2.4. Influencia de la edad y el sexo de los progenitores sobre las mutaciones	27
3. Las mutaciones en el contexto de la genética	34
3.1. Breve reseña sobre la evolución de la selección animal	35
3.2. La variabilidad mutacional <i>de novo</i> en las poblaciones	40
Bibliografía	44
Objetivos	59
Materiales y Métodos	63
1. Modelo analítico desarrollado	65
1.1. Parametrización del modelo de análisis	65
1.2. Desarrollo Bayesiano	67
1.3. Resolución del modelo Bayesiano	68
1.4. Comparación de modelos	69
2. Poblaciones y datos analizados	72
2.1. Ratones isogénicos	72
2.2. Vacuno cárnico	73
3. Modelos operacionales	76
Bibliografía	77
Resultados	81
1. Desarrollo metodológico	83
2. Análisis sobre especies domésticas	85
2.1. Ratones C57BL/6J	85
2.2. Terneros de raza Bruna de los Pirineos	88
Bibliografía	92
Discusión	95
1. Metodología y contexto de análisis	97
2. Varianza mutacional en ratones isogénicos C57BL/6J	103
3. Varianza mutacional en terneros de raza Bruna de los Pirineos	107
4. Implicaciones de la varianza mutacional	110
Bibliografía	113
Conclusiones	117

CAPÍTULO I
INTRODUCCIÓN

1. LOS ÁCIDOS NUCLEICOS EN EL CONTEXTO EVOLUTIVO

1.1. Definición, tipos y características

La información genética de todo organismo viviente se transmite a la descendencia a través de moléculas denominadas “ácidos nucleicos”. Su descubrimiento data de la segunda mitad del siglo XIX, cuando J. F. Miescher aisló una sustancia ácida rica en fosfatos a partir de núcleos de glóbulos blancos (Dahm, 2008). Estas moléculas se definen como polímeros formados por unidades de nucleótidos que a su vez, están compuestos por un hidrato de carbono o pentosa, una base nitrogenada y un grupo fosfato. Dependiendo de la naturaleza de la pentosa se pueden distinguir dos tipos de ácidos nucleicos, (1) el ADN o ácido desoxirribonucleico que contiene 2-desoxi-D-ribosa, y (2) el ARN o ácido ribonucleico que contiene la D-ribosa. Las bases nitrogenadas (*i.e.*, nucleótidos) que conforman los ácidos nucleicos pueden ser monocíclicas (pirimidinas) o bicíclicas (purinas). Las bases que intervienen en la formación de ADN se denominan adenina (A), citosina (C), guanina (G) y timina (T), mientras que el ARN contiene las tres primeras, más el uracilo (U) en lugar de la ya mencionada timina (Murray et al., 2012).

Con posterioridad a los estudios realizados por J. F. Miescher que llevaron a la identificación de un material que llamó nucleína, en 1869, se realizaron estudios que fueron caracterizando progresivamente la naturaleza química de los ácidos nucleicos, así como su importancia biológica (Kossel y Steudel, 1903). Por un lado, O. T. Avery, M. C. MacLeod y M. McCarthy demostraron que la introducción de ADN purificado en bacterias hacía que éstas cambiaran su fenotipo y, por lo tanto, que el ADN era la molécula transmisora de la información genética (Avery et al., 1944). Poco tiempo después, E. D. Chargaff estableció que las proporciones de cada tipo de base nitrogenada en los ácidos nucleicos eran idénticas para las purinas (*i.e.*, A y T) y para las pirimidinas (*i.e.*, C y G; Chargaff, 1952; Chargaff et al., 1952).

Estos descubrimientos culminaron en 1953, cuando J. D. Watson, F. H. Crick, R. Franklin y M. H. F. Wilkins (además de otros colaboradores), describieron la estructura tridimensional del ADN (*i.e.*, la doble hélice) a través de varios trabajos publicados en la revista Nature (Franklin y Gosling, 1953; Watson y Crick, 1953a,b; Wilkins et al., 1953).

El modelo de la doble hélice de ADN de Watson y Crick (1953a,b) establece una estructura compuesta por dos cadenas complementarias que forman una doble hélice. Esta conformación se sustenta gracias a la existencia de enlaces con tres puentes de hidrógeno para la unión entre G y C, y dos puentes de hidrógeno para la unión entre A y T, brindando la estabilidad necesaria a la molécula. Al mismo tiempo, las cadenas de nucleótidos de la molécula de ADN adoptan una configuración helicoidal donde las bases nitrogenadas ocupan el interior de la estructura y se disponen perpendicularmente a las cadenas laterales. Estas cadenas de nucleótidos tienen direccionalidad y polaridad concretas, siendo las dos cadenas antiparalelas (*i.e.*, con polaridad contraria); cada una está formada por una cadena de nucleótidos unidos por enlaces fosfodiéster, y las bases nitrogenadas están unidas a las pentosas mediante enlaces glucosídicos (Novo, 2007).

1.2. Estructura y función del ADN

El descubrimiento de que la información genética estaba codificada a lo largo de una molécula polimérica compuesta de sólo cuatro tipos de unidades monoméricas, fue uno de los principales logros científicos del siglo XX. Esta molécula polimérica, el ADN (o ARN en algunas etapas de la fisiología celular), representa la base química de la herencia y se organiza en genes, las unidades fundamentales y funcionales de la información genética. Un gen se puede entender como la sección de ADN que codifica para un producto funcional, siendo este generalmente una proteína, aunque también se puede tratar de ARN mensajero, ácido

ribonucleico ribosómico (ARNr), o ARN de transferencia (ARNt; Pearson 2006). Además, se han descrito también múltiples variantes de ARN no codificante a partir del descubrimiento llevado a cabo por el laboratorio de V. Ambros hace ya casi 25 años (Lee et al., 1993). Tal como resume Taft et al. (2010), las distintas categorías de ARN no codificante incluye (1) RNA no codificante largos (*long non-coding RNA*), (2) RNA pequeños de interferencia (*small interfering RNA*), (3) microRNA, RNA de interacción (*interacting RNA*), (4) RNA asociados a promotores (*promoter-associated RNA*), y (5) RNA nucleares pequeños (*small nucleolar RNA*).

Debe tenerse en cuenta que los genes no funcionan de modo autónomo; su replicación y función están controladas por diversos productos genéticos, a menudo en colaboración con componentes de diversas vías de transducción de señal, generando redes complejas de interacción. El conocimiento de la estructura y función de los ácidos nucleicos es esencial para entender los aspectos genéticos de la fisiopatología, así como la base genética de los distintos fenotipos de interés, desde procesos patológicos en humanos hasta caracteres productivos en animales domésticos (Murray et al., 2012).

La información contenida en el ADN (el código genético) reside en la secuencia en la que están ordenados los nucleótidos. El polímero de ADN, como se ha dicho anteriormente, posee una polaridad; un extremo tiene un grupo 5'-hidroxilo o fosfato terminal, mientras que el otro tiene un 3'-fosfato o hidroxilo terminal (Murray et al., 2012). La importancia de esta polaridad quedará de manifiesto al discutir los mecanismos de replicación y transcripción del ADN. De hecho, dado que la información genética reside en el orden de las unidades monoméricas dentro de los polímeros, resulta imprescindible que haya un mecanismo eficiente para reproducir o replicar el código genético, con un alto grado de fidelidad. Debemos tener en cuenta que la información genética almacenada en la secuencia de nucleótidos del ADN tiene dos propósitos básicos, (1) es la fuente de información básica para la síntesis de todas las moléculas de proteína (y otras macro-moléculas) imprescindibles para el

correcto funcionamiento de la célula y el organismo, y (2) representa el mecanismo básico de herencia a partir de la secuencia genética contenida en el genoma de los progenitores (Mendel, 1866). Ambas funciones requieren que la molécula de ADN sirva como una plantilla, en el primer caso para la transcripción de la información hacia el ARN, y en el segundo para la duplicación (o replicación) de la información hacia moléculas de ADN hijas, pero en cualquier caso, ambos procedimientos deberían realizarse sin originar modificaciones en la secuencia. Cuando se originan dichas modificaciones, estaríamos hablando de mutaciones, las cuales representan un fenómeno biológico complejo, aunque relevante, sobre el que se centra esta tesis doctoral.

La cromatina es el material genético organizado en los núcleos de las células de eucariotas y consta de (1) moléculas de ADN bicatenario muy largas, (2) una masa casi igual de proteínas básicas con un tamaño molecular pequeño (*i.e.*, histonas), (3) una cantidad menor de proteínas no histona (la mayor parte de las cuales son ácidas y de mayor tamaño que las histonas), y (4) una pequeña cantidad de ARN (Alberts et al., 2002). Las proteínas no histona incluyen enzimas involucradas en la replicación y reparación del ADN, así como son también proteínas participantes en la síntesis, el procesamiento y el transporte del ARN hacia el citoplasma. La hélice del ADN se estructura en unidades físicas conocidos como cromosomas, y la doble cadena de nucleótidos de cada cromosoma tiene una longitud que es miles de veces el diámetro del núcleo de la célula (Murray et al., 2012). Es precisamente por esto que uno de los propósitos de las moléculas que comprenden la cromatina, en especial las histonas, es condensar el ADN y formar los cromosomas. Sin embargo, es importante notar que las histonas también pueden tener una participación fundamental en la regulación de los genes; en efecto, las histonas contribuyen de manera importante en muchas de las transacciones moleculares que implican ADN (Alberts et al., 2002). En cualquier caso, las histonas permiten la formación de los cromosomas, que son una forma especialmente condensada y dispuesta de la cromatina durante buena parte del ciclo celular. Esta condensación se

origina por un enrollamiento en forma de espiral de las diferentes fibras de la cromatina. Al principio el término cromosoma sólo se refería a las estructuras densas que se teñían de tono oscuro mediante tinciones específicas, observables en el interior de las células eucariotas durante la meiosis o la mitosis (estos conceptos se refieren a distintos procedimientos de división celular, y se discutirán más adelante). Sin embargo, este término se utiliza también en la actualidad para describir a las moléculas de ADN de las células procariotas. Cada cromosoma eucariótico consta de una sola molécula lineal de ADN que forma un complejo con histonas. Paralelamente, se emplea el término cromatina para describir el complejo formado por ADN e histonas sin llegar a condensarse en forma de cromosomas (también se unen a la cromatina otros tipos de proteínas no histonas; algunos ejemplos son las enzimas que participan en la replicación y en la reparación del ADN, y una gran cantidad de factores de transcripción de diferentes tipos, tanto activadores como represores, entre otros; McKee et al., 2015).

Según Murray et al. (2012), existen evidencias de que la organización nuclear y la estructura de la cromatina están involucradas en la determinación de la regulación y el inicio de la síntesis de ADN. El índice de polimerización en células eucarióticas, que tienen cromatina y nucleosomas, es más lento que el que se observa en células procariotas, que carecen de cromosomas canónicos. También está claro que la estructura de cromatina debe volver a estructurarse luego de la replicación. El ADN recién replicado se asocia con rapidez para formar los nucleosomas (*i.e.*, octámero de proteínas histonas y unos 147 pares de bases nitrogenadas), ya que los octámeros de histonas preexistentes y recién sintetizados se distribuyen al azar hacia cada brazo de la horquilla de replicación del ADN. Estas reacciones se facilitan mediante la participación de proteínas específicas (*e.g.*, chaperonas) que trabajan conjuntamente dentro del contexto de los complejos remodeladores de la cromatina (Murray et al., 2012).

1.3. Replicación y transcripción

Las proteínas, como principales elementos estructurales y funcionales de las células, se determinan a partir de la secuencia de nucleótidos en las regiones codificantes del ADN. La secuencia de nucleótidos determina el tipo de aminoácido y el orden en el que se suceden durante el proceso de síntesis proteica. El proceso por el cual la información contenida en el ADN es descodificada y traducida para dar lugar a la síntesis de proteínas se conoce como expresión génica y comprende varios pasos de descodificación que son, en líneas generales, (1) la transcripción y (2) la traducción.

El proceso de transcripción consiste en la síntesis de una cadena de ARN por acción enzimática a partir de una molécula de ADN y utilizando nucleótidos libres disponibles en el núcleo celular. En este proceso, actúan enzimas de tipo polimerasa que se deslizan a lo largo de la cadena molde de ADN en un sentido específico, transcribiendo a ARN de manera cíclica y produciendo el ARN mensajero (ARNm). En eucariotas existen varios tipos de enzimas polimerasas en función del tipo de gen que transcriben (Novo, 2007). De hecho, en organismos eucariotas el proceso de transcripción es sumamente complejo y abarca una serie de pasos que incluyen la iniciación, la elongación y la terminación, sin contar otros eventos como la preiniciación o la disgregación del promotor, justo antes y después de la iniciación, que podrían incluso considerarse como pasos adicionales del proceso. Todos ellos llevan implícitos una serie de eventos y reacciones bioquímicas donde intervienen diferentes factores y complejos enzimáticos, sobre los cuales no es preciso ahondar en estos momentos para los fines de la presente revisión.

Resulta importante señalar que de las cuatro clases principales de ARN, el ARN ribosómico (ARNr), el ARN mensajero (ARNm) y el ARN de transferencia (ARNt) participan activamente en la síntesis de proteínas mientras que los ARN pequeños no codificantes intervienen en otros procesos, incluyendo el acoplamiento

del ARNm con el ribosoma o la modulación de la expresión del gen. Estas diferentes clases de ARN se distinguen entre sí por su estabilidad, abundancia, diversidad y funcionalidad dentro de la célula. El ARNm es el tipo de mayor abundancia, tamaño y estabilidad, y además resultan más heterogéneos. Se sintetiza en el núcleo de la célula y posteriormente se transportan al citoplasma, atravesando la membrana nuclear. De hecho, funcionan como mensajeros que transmiten la secuencia codificante de un gen hacia la maquinaria sintetizadora de proteína, donde cada ARNm sirve como una plantilla sobre la que se polimeriza la secuencia de aminoácidos para formar una molécula de proteína específica, el producto final de gen (Murray et al., 2012). El ARNt, varía desde 74 hasta 95 nucleótidos, y también se generan por el procesamiento nuclear de una molécula precursora. Las moléculas de ARNt sirven como adaptadoras para la traducción de la información en la secuencia de nucleótidos del ARNm hacia aminoácidos específicos (ver apartado siguiente). Hay al menos 20 especies distintas de moléculas de ARNt en cada célula, y por lo menos una (y a menudo varias) corresponde a cada uno de los 20 aminoácidos requeridos para la síntesis de proteína (Novo, 2007). Un ribosoma es una estructura nucleoproteínica citoplásmica que actúa como la maquinaria para la síntesis de proteínas a partir de las plantillas de ARNm. En los ribosomas, las moléculas de ARNm y ARNt interactúan para traducir el ARNm hacia una molécula de proteína con secuencia específica según la información contenida en el gen (Novo 2007).

1.4. Traducción: Síntesis de proteínas

Por simplicidad y tal como se ha comentado anteriormente, las letras A, G, T y C corresponden a los nucleótidos que conforman el ADN (más U en lugar de T para el ARN). Estos nucleótidos se organizan secuencialmente formando los genes que codifican para las distintas proteínas. Para el proceso de traducción del material

genético a proteínas, los nucleótidos del ARNm se leen de tres en tres, y cada bloque de tres nucleótidos se denomina codón. El conjunto de todos los codones posibles constituyen el Código Genético (Shu 2017). Antes de que se elucidara el código genético, era imposible entender la síntesis de proteína, o explicar el efecto y las consecuencias de las mutaciones. Además, el código proporciona el punto de partida para explicar la manera en la cual las alteraciones del código genético (*i.e.*, mutaciones) pueden originar cambios también a nivel de las proteínas, y estos, a su vez, causar alteraciones morfológicas, fisiológicas o metabólicas en el organismo (Novo, 2007).

En términos generales, la información genética de la secuencia de nucleótidos del ADN se transcribe en el núcleo de la célula en forma de una secuencia de nucleótidos de ARN, conociéndose como ARNm (*i.e.*, transcripción). Evidentemente, la secuencia de nucleótidos del ARN mensajero es complementaria a la secuencia de nucleótido de la cadena original de ADN para el gen en cuestión. La célula posee la maquinaria necesaria para traducir con exactitud y eficacia la información desde la secuencia de nucleótidos de un ARNm hacia la secuencia de aminoácidos de la proteína específica correspondiente. No obstante, desde el principio quedó de manifiesto que las moléculas de ARNm mismas carecían de afinidad por los aminoácidos y, por ende, que la traducción de la información en la secuencia del ARNm hacia la secuencia de aminoácidos de la proteína requería una estructura adaptadora intermedia, que debía reconocer una secuencia de nucleótido específica por un lado, así como un aminoácido específico por el otro. Esta estructura es el ribosoma, un complejo macromolecular constituido por proteínas y ARN ribosómico, que permite dirigir a un aminoácido específico hacia la posición secuencial apropiada de una proteína durante su síntesis, según lo dicta la secuencia de ARNm específica. De hecho, no hay contacto directo real entre los aminoácidos y el ARNm, aunque unos y otros están íntimamente relacionados.

Los ARNm de eucariotas tienen una característica muy importante, no son

codificantes en su totalidad (*i.e.*, no toda la secuencia se traduce a proteína), desde el principio al fin, sino que las regiones codificantes están interrumpidas por otras regiones no codificantes. En otras palabras, existen regiones codificantes llamadas exones que alternan con otras regiones no-codificantes llamadas intrones. Debido a esta configuración, se requiere de un paso previo antes de la síntesis de proteínas que consiste en eliminar los intrones y pegar los exones para formar un ARNm que pueda ser traducido íntegramente, sin interrupciones y con un sentido biológico real. Este proceso de corte y eliminación de intrones, así como el de unión de los exones, se denomina en inglés *splicing*, y fue descrito por primera vez hace ya cuarenta años (Berget et al., 1977; Chow et al., 1977).

Resulta fácil calcular el número posible de combinaciones de tres letras que se pueden formarse a partir de un alfabeto de cuatro letras (A, C, G y T), 43, es decir, 64 tríos (o codones) distintos. Sin embargo, únicamente se han descrito 20 aminoácidos esenciales distintos en las proteínas, por lo que sólo resultarían necesarios 20 codones, y los 44 restantes serían redundantes. Este exceso de codones posibles se interpreta como un fenómeno natural por el que se considera que el código genético está degenerado, ya que varios codones codifican para el mismo aminoácido. A modo de ejemplo, el aminoácido leucina está codificada por seis codones diferentes (UUA, UUG, CUU, CUC, CUA, CUG), mientras que el triptófano únicamente se codifica a partir del codón UGG. Esto se origina debido a un fenómeno conocido como *wobble pairing*, por el que el tercer nucleótido del codón puede no emparejarse con su homólogo de manera estricta, además que en los RNAt puede aparecer un quinto aminoácido posible, la inosina (Campbell y Reece, 2011).

1.5. El ciclo celular

Aunque la presente tesis doctoral no se centra en los procesos biológicos

específicos en los que interviene el material genético (ADN o ARN), si lo hace sobre las consecuencias de los mismos en forma de errores incluidos en la secuencia, las mutaciones. Para ello, resulta importante comentar el ciclo celular, así como los distintos puntos del mismo en el que se pueden originar (y transmitir) dichas mutaciones. Desde un punto de vista generalista, la reproducción celular consiste en la formación de dos o más células hijas que deberían tener la misma información genética que la célula madre. Par a ello resulta imprescindible que se duplique el ADN y, posteriormente, que el núcleo se divida en dos mediante un proceso llamado cariocinesis; este último, generalmente va seguido de una división del citoplasma o citocinesis.

El ciclo celular es el período de biosíntesis y crecimiento de las células, y consta de dos períodos o fases principales, (1) la interfase, que es la organización de la célula para la autoconservación y ocupa aproximadamente el 90% del ciclo, y (2) la mitosis (o fase M), que consiste en los mecanismos necesarios para la duplicación de la célula y el reparto equitativo del material genético.

1.5.1. Mitosis

Un aspecto básico de la mitosis para la posible aparición de mutaciones (Kodym y Afza, 2003), es la duplicación del material genético, con el objetivo que cada célula resultante disponga de una copia idéntica de cada cromosoma y, en consecuencia, de toda la secuencia de ADN. Esto se produce durante la interfase, aprovechando que en este momento el material genético no se encuentra condensado formando los cromosomas sino de dispuesto de manera laxa en el núcleo celular, en forma de cromatina. En eucariotas, la replicación del material genético comienza en múltiples sitios de una misma molécula de ADN a la vez, llamados orígenes de replicación. Dada la velocidad de copia de las polimerasas, una sola molécula de ADN polimerasa tardaría varios días en replicar un cromosoma completo, por lo que

de hecho, la replicación de la cromatina se lleva a cabo en varios miles de orígenes de replicación a la vez (Murray et al., 2012). En cada origen de replicación se forma una burbuja de replicación. Mediante la acción de unas enzimas conocidas como helicasas, se abre la doble hélice de ADN para permitir la unión de las enzimas que llevarán a cabo la síntesis, que son las polimerasas de ADN. En cada burbuja se forman dos horquillas de replicación, apuntando en direcciones contrarias, y duplicando cada una de las dos cadenas de nucleótido a la vez; una cadena se duplica de manera continua, y la otra, por fragmentos discontinuos, conocidos como fragmentos de Okazaki (Okazaki et al., 1968; Sugimoto et al., 1968).

La cromatina es una estructura dinámica cuyo grado de empaquetamiento es máximo durante la mitosis y, por ello, en esta fase del ciclo celular se puede ver en una forma especialmente condensada que da lugar a los cromosomas. Esta condensación se produce porque la fibra de cromatina se enrolla en forma de espiral, proporcionando un grado de empaquetamiento unas cinco veces mayor al observado durante la interfase (fase en que la célula experimenta sus funciones celulares de mantenimiento celular). La mitosis es el proceso de división de las células somáticas, fundamental en la proliferación celular que tiene lugar durante el desarrollo embrionario, el crecimiento y el mantenimiento de los tejidos. Supone una reorganización drástica de todos los componentes celulares, pero muy especialmente de los cromosomas, cuya segregación a cada una de las células hijas debe ser muy precisa y estar finamente regulada y coordinada con la separación física de las nuevas células (*i.e.*, citoquinesis). Durante la mitosis, la maquinaria celular se especializa en llevar a cabo los distintos procesos que tienen lugar en la célula: condensación de la cromatina, formación del huso mitótico, segregación de los componentes, y fisión celular.

La primera fase de la mitosis (*i.e.*, profase), comienza con la condensación de la cromatina, la ruptura de la membrana nuclear y la creación del huso mitótico (*i.e.*, conjunto de microtúbulos que brotan de estructuras diametralmente opuestas en la

célula y conocidas como centriolos). Es importante recordar que la cromatina ha sido replicada durante la interfase previa, por lo que cada cromosoma está ahora formado por dos cromátidas hermanas idénticas (a excepción de que haya podido originar algún error durante la duplicación, lo que significaría la aparición de una nueva mutación en la secuencia de ADN). Los microtúbulos se unen a los centrómeros de cada cromosoma, y comienzan a arrastrar los cromosomas hacia el plano ecuatorial del huso. En la fase siguiente (*i.e.*, metafase) cada cromosoma está unido a microtúbulos procedentes de los dos centrómeros opuestos de la célula, de modo que todos los cromosomas están en el ecuador del huso mitótico sometidos a fuerzas tensionales opuestas. La metafase va seguida por la anafase, en la que tiene lugar la segregación de las cromátidas hermanas de cada cromosoma hacia polos opuestos de la célula. De hecho, en esta fase las proteínas que mantenían unidas ambas cromátidas hermanas (*i.e.*, cohesinas) son cortadas, a la vez que los microtúbulos anclados a cada centrómero traccionan de la cromátida correspondiente. La separación simultánea de los pares de cromátidas hermanas en la transición metafase-anafase es un momento crucial del ciclo celular, y por tanto está finamente regulado. Por ejemplo, es crítico que la cohesión se pierda en el momento adecuado, para que cada cromátida pueda migrar a una célula hija sin errores. La última fase de la mitosis es la telofase, en la que los cromosomas vuelven a descondensarse y se forma la envoltura nuclear alrededor de cada uno de los nuevos núcleos que se han formado en cada polo de la célula. Terminada la mitosis, el proceso de división celular se completará con la citocinesis, en la que los componentes celulares se reordenan y se reorganiza el citoesqueleto para facilitar la división física de la célula en dos células hijas (Alberts et al., 2002; Novo, 2007).

1.5.2. Meiosis

En organismos de reproducción sexual, la formación de gametos (óvulos o

espermatozoides) implica un modo de división celular diferente al de las células somáticas. Los gametos deben poseer una sola copia de cada cromosoma (*i.e.*, son haploides), de modo que el cigoto resultante de la unión del gameto masculino y el femenino durante la fertilización sea diploide (*i.e.*, tenga dos copias de cada cromosoma). El modo de división celular que permite la creación de células haploides se denomina meiosis, y tiene además otras características muy importantes para la transmisión de la variabilidad genética.

En ambas líneas germinales, masculina o femenina, la gametogénesis comienza cuando los gonocitos experimentan una serie de divisiones mitóticas típicas en las que las células van madurando hasta llegar a formar las “gonias” (espermatogonias u oogonias, respectivamente) y los gametocitos primarios, que son células diploides en todos los casos. A partir de este momento, las células replican su ADN, tal como sucedía en la mitosis, pero a continuación entran en meiosis, en vez de dividirse nuevamente por mitosis. En términos generales, la meiosis incluye en realidad dos divisiones celulares distintas y sucesivas, por lo que suele separarse en dos fases llamadas meiosis I (primera división meiótica) y meiosis II (segunda división meiótica). La meiosis I es la más importante de las dos, y comienza con una profase larga, complicada y claramente distinta de la profase de la mitosis. Esta profase de la meiosis I (llamada, por tanto, profase I) comienza con la condensación de la cromatina que, no lo olvidemos, se había replicado anteriormente, exactamente igual que antes de cualquier división por mitosis. Ambas cromátidas están íntimamente unidas formando un único filamento, que se une por sus extremos a la membrana nuclear durante la primera parte de la profase I, llamada leptoteno. En la siguiente etapa, llamada cigoteno, los cromosomas homólogos se emparejan longitudinalmente con gran exactitud, de manera que las secuencias de cada uno de los dos homólogos quedan perfectamente alineadas. La unión entre los cromosomas de cada par se hace progresivamente más fuerte, dando lugar a un fenómeno que se conoce como sinapsis (la sinapsis se mantiene gracias a la formación del complejo sinaptonémico). La profase I continúa con la siguiente etapa, paquiteno, en la que los

cromosomas se condensan todavía más y cada par de homólogos aparece como un bivalente (una estructura formada por dos cromosomas) o tétrada (por estar formada por cuatro cromátidas). De todas formas, el evento más importante de la etapa de paquiteno es el intercambio de material genético entre cromosomas homólogos, a través del proceso de la recombinación (Holliday, 1964). En la siguiente etapa, la de diploteno, desaparece el complejo sinaptonémico y los cromosomas homólogos comienzan a separarse. Recordemos que hasta este momento la célula es diploide y con dos cromátidas por cromosoma. Durante la separación de los cromosomas homólogos, en el diploteno, las dos cromátidas de cada cromosoma permanecen todavía unidas entre sí. Sin embargo, debido al intercambio de material genético precedente, la separación pone de manifiesto los puntos en los que se produjo una recombinación, ya que esas regiones cruzadas forman puentes que mantienen unidos a los dos homólogos de cada par e impiden la separación total. Estos puntos se denominan quiasmas, y han de resolverse de algún modo para que la separación de los cromosomas pueda completarse (existen distintos modelos distintos de resolución). La profase I termina con una etapa corta denominada diaquinesis, en la que la cromatina alcanza su grado máximo de condensación, y los cromosomas aparecen más cortos y gruesos (McKee et al., 2015).

Tras la profase I, los espermatoцитos primarios entran ahora en metafase I, en la cual desaparece la membrana nuclear, se forma el huso y los microtúbulos traccionan por los quinetocoros. Esto hace que los bivalentes, que llevan los cromosomas parcialmente separados (unidos únicamente por los quiasmas), se orienten en el ecuador de la célula. Debido a esta configuración, los cromosomas homólogos se sitúan en la placa metafásica con los centrómeros apuntando cada uno hacia uno de los polos de la célula. En la anafase I, la tracción de los microtúbulos y la activación del complejo promotor de la anafase permiten la separación total de cada cromosoma homólogo hacia uno de los polos, fenómeno llamado disyunción. Finalmente, en la telofase I se forman ya dos grupos de cromosomas en cada polo de la célula; cada uno de estos grupos consta de un número haploide de cromosomas

que poseen dos cromátidas cada uno. La meiosis I termina con la citoquinesis, la división física en dos células hijas. Esta última fase es diferente en la línea germinal masculina y femenina, porque así como las células hijas del espermatocito primario son iguales, en el caso de la línea femenina una de las dos células hijas se lleva casi todo el citoplasma mientras que la otra es mucho menor y carece de funcionalidad. Por tanto, en la espermatogénesis cada espermatocito primario da lugar a dos espermatocitos secundarios, mientras en la ovogénesis, cada oocito primario da lugar a un oocito secundario y a un corpúsculo polar de primer orden (Alberts et al., 2002; Novo, 2007).

La segunda división meiótica, o meiosis II, es prácticamente igual a una mitosis. La única diferencia estriba en que la célula que se divide es haploide, mientras que en el caso de la mitosis la célula era diploide. Por tanto, esta división celular va a separar las dos cromátidas que componen cada cromosoma de un gametocito secundario, para generar gametos con un número haploide de cromosomas, cada uno de los cuales está formado por una sola cromátida del cromosoma. Como se comentaba al principio, esto asegura que la fecundación da lugar a un cigoto diploide, al recibir un complemento cromosómico de cada gameto paterno.

El modo en que se completa la meiosis es también diferente en ambos sexos. Mientras la espermatogénesis discurre de un modo continuo desde que se alcanza la madurez sexual, la ovogénesis se detiene en profase I, de manera que un alto número de oocitos primarios permanecen en un estado prolongado de diploteno, llamado dictioteno. Estos oocitos sólo proseguirán su maduración al llegar a la madurez sexual de la hembra. Con cada ciclo reproductivo, uno o varios oocitos culminan la meiosis I y completan también la meiosis II para, cada uno de ellos, dar lugar a un oocito secundario (óvulo) y un segundo corpúsculo polar que nuevamente carece de funcionalidad biológica o reproductiva, y se desecha sin mayores consideraciones (Murray et al., 2012).

La mayor importancia de la meiosis reside en que es uno de los principales mecanismos de creación de variabilidad genética. Esto sucede a dos niveles, primero por la segregación aleatoria de los cromosomas de cada progenitor en los gametos, y en segundo lugar, por el intercambio de material genético que tiene lugar durante los procesos de recombinación. Respecto a lo primero, es fácil comprender que cada división meiótica producirá gametos con distintas combinaciones de los cromosomas que estaban en la célula diploide original, ya que la segregación se produce al azar. Para entender esto es útil imaginarse todos los cromosomas de una célula diploide como la suma de dos conjuntos haploides, el paterno y el materno. Así, cada pareja de cromosomas homólogos consta de un homólogo de origen paterno y otro de origen materno. Cuando los homólogos se separan en la meiosis, resulta poco probable que pasen todos los de origen paterno a una de las células hijas, y todos los de origen materno a la otra; por el contrario, se distribuyen aleatoriamente. Como los homólogos pueden tener pequeñas diferencias en los alelos presentes para algunos genes, el resultado es que cada gameto lleva una combinación única de cromosomas de origen paterno y materno mezclados. La variabilidad que se puede generar por este sencillo mecanismo es altísima. Por lo que respecta al segundo tipo de variabilidad introducida durante la meiosis, en la especie humana se ha calculado que se producen unas 55 recombinaciones por célula, dado que este es el número medio de quiasmas que se observan (Coop y Przeworski 2007). Esto supone una media de al menos dos recombinaciones por par cromosómico, incluidos los cromosomas sexuales. La presencia de estos quiasmas es importante para mantener los bivalentes unidos hasta la anafase I, y para asegurar que la disyunción se produce de modo correcto. Pero además, es probable que el número total de eventos de recombinación sea todavía mayor al número de quiasmas, porque no todas las recombinaciones producen un sobrecruzamiento. Por tanto, el grado de intercambio de material genético entre cromáticas de cromosomas homólogos es bastante alto, y hace que los cromosomas que finalmente llegan a un gameto sean ligeramente distintos a los cromosomas que estaban presentes en la célula inicial. Como resultado de esto, la

variabilidad genética generada durante la meiosis es alta y la probabilidad de que dos individuos tengan dos descendientes genéticamente idénticos es prácticamente nula. Esto es precisamente lo que persigue la reproducción sexual, generar descendientes distintos a sus progenitores y distintos entre sí, aunque las diferencias no sean absolutas (Novo, 2007).

Uno de los aspectos más importantes en el campo de la genética es la caracterización de los mecanismos moleculares por los que se produce intercambio de material genético entre cromátidas durante la meiosis. Distintos modelos han intentado explicar cómo dos moléculas de ADN homólogas pueden entrecruzarse e intercambiar material genético. El paso inicial común a todos los modelos propuestos es el alineamiento preciso de ambas moléculas, justamente debido a que son cadenas de ADN homólogas, es decir, pertenecientes a cromosomas homólogos y por tanto con (casi) la misma secuencia de nucleótidos. En realidad, hoy sabemos que ambas moléculas no son absolutamente idénticas, ya que pueden existir diferencias alélicas en su secuencia. El alineamiento de dos secuencias homólogas es un requisito imprescindible para que se inicie el proceso de recombinación, que se denomina por eso recombinación homóloga. Dicho alineamiento, como es lógico, se produce durante la profase I de la meiosis merced al complejo sinaptonémico que mantiene estrechamente unidos los cromosomas homólogos en toda su longitud (Alberts et al., 2002).

2. LA VARIABILIDAD GENÉTICA Y LAS MUTACIONES

2.1. Variabilidad genética

La variabilidad observable en el fenotipo de los individuos de una población determinada se atribuye en gran medida a la variabilidad genética, sobre la cual se añade aquella producida por los efectos ambientales. La variación genética se debe a la presencia de ciertos cambios genéticos generalmente heredables (de una célula a sus células hijas) y transmisibles a la siguiente generación si se ve afectada la línea germinal paterna o materna (si solamente se ven afectadas las células somáticas, estos cambios no serán transmitidos). Como es sabido, un gen puede presentarse en diferentes formas debido a variaciones en la secuencia de nucleótidos; cada forma alternativa de un gen se denomina alelo y la proporción de copias de dicho alelo en una población se conoce como frecuencia alélica. Cuando un gen presenta al menos dos formas alélicas distintas y la frecuencia alélica del alelo más raro es del 1% o mayor, ese gen se considera polimórfico y la variación se define como polimorfismo. La variabilidad entre individuos garantiza la posibilidad de adaptación de una especie a condiciones ambientales cambiantes, pero también puede ser motivo de alteraciones indeseadas que pueden conllevar a presencia de desórdenes genéticos como pudieran ser enfermedades (Swallow, 2003; Futreal et al., 2004; Watson et al., 2013).

La variación entre individuos, tal como fue señalado anteriormente, se genera principalmente durante el proceso de reproducción sexual, por recombinación meiótica. Además, durante la vida de un individuo se van introduciendo cambios originados como consecuencia de nuevas mutaciones en el genoma de sus células, aunque sólo un pequeño porcentaje de estos cambios afectan a las células germinales y pueden ser transmitidos al genoma de las nuevas generaciones. La tasa final de nuevas mutaciones resulta del equilibrio entre los fenómenos de mutación y reparación de las mutaciones. Se sabe que en seres humanos, la ADN polimerasa

produce aproximadamente un error cada 10.000 nucleótidos incorporados durante el proceso de transcripción de la síntesis proteica (Eyre-Walker y Keightley 2007) y está dotada de un sistema adicional de “edición” que reduce significativamente la tasa de nuevas mutaciones, generando finalmente un error cada aproximadamente 10.000.000 nucleótidos (Alberts et al., 2002). Además, los sistemas genómicos de detección y reparación de errores corrigen muchas de las mutaciones primarias, consiguiendo que, al final, en la replicación del ADN celular solo se origine un error cada aproximadamente 10.000.000.000 nucleótidos. De hecho, se estima que debería aparecer una nueva mutación cada 300 divisiones celulares, asumiendo un total de 30.000 genes para el caso específico del genoma humano (Novo, 2007). Esta es una tasa de mutación extremadamente baja, lo que implica la existencia de sistemas celulares de reparación del ADN que se ocupan de corregir la inmensa mayoría de las mutaciones introducidas en el genoma. Además, muchas de estas mutaciones serán silenciosas porque no afectarán a ADN codificante; otras, sufrirán una fuerte presión selectiva y su frecuencia alélica será muy baja (o llegarán a desaparecer). Como se sugiere, la presencia de polimorfismos y mutaciones en las poblaciones serán el resultado de la relación entre mutación-reparación-selección.

No obstante, y aunque la tasa de mutación se sugiera muy baja, su impacto en las poblaciones queda fuera de toda duda, ya que ellas son el punto de partida para la generación de variabilidad genética *de novo* (Hill, 1982a,b; Casellas y Medrano, 2008), sustentan la respuesta a la selección a largo plazo (Keightley, 1998; Hill, 2007), y pueden originar cambios fenotípicos drásticos tanto a nivel fisiológico (Bradford y Famula, 1984; Guenther et al., 2014) como patológico (Nishi y Nanjo, 2011; Straniero et al., 2015).

2.2. Mutaciones

Una mutación es un cambio, en general irreversible, de una parte del genoma

de un individuo. Este cambio, que por otro lado puede no ser observado inmediatamente a nivel del fenotipo, se transmite como un nuevo rasgo hereditario si afecta a las células germinales. Las mutaciones son frecuentes en mamíferos y tienen un interés considerable para los genetistas, además de su intrínseco valor evolutivo, por ser generadoras de diversidad genética. Todos los seres vivos, sin excepción, sufren mutaciones en sus genomas y este proceso es fundamentalmente aleatorio. Las mutaciones pueden presentarse sobre células somáticas o germinales, en células embrionarias o adultas, en cada caso con consecuencias potencialmente diferentes. Pueden afectar tanto secuencias que codifican para proteínas como secuencia reguladoras o repetitivas del genoma. A su vez, las mutaciones acostumbran originar tanto ventajas como inconvenientes, imprevisibles *a priori* dada la aleatoriedad intrínseca del proceso mutacional. Por un lado, la aparición de mutaciones permite a los individuos adquirir nuevos polimorfismos en su genoma, incrementando la variabilidad genética de la población y, en consecuencia, evolucionar; no obstante, a menudo las mutaciones pueden implicar efectos deletéreos o patológicos, degenerando incluso hasta el nivel de originar enfermedades hereditarias (Benavides y Guénet 2003).

Durante el desarrollo embrionario temprano, las células del organismo se multiplican a una velocidad muy rápida y, por lo tanto, se incrementa la probabilidad de aparición de nuevas mutaciones. No obstante, su aparición no tiene por qué conllevar un impacto directo sobre el individuo, ni su transmisión sistemática a las nuevas generaciones. De todas las mutaciones que se generan, algunas provocan directamente la muerte de la célula afectada, en consecuencia, se eliminan del organismo. En otros casos, las mutaciones pueden impedir que las células se diferencien hacia un tejido en particular y son, por lo tanto, fuertemente contra-seleccionadas. Además, algunas mutaciones no afectan ni las funciones ni la supervivencia de la célula, o incluso no provocan ningún cambio a nivel de la secuencia de la proteína, por lo que se consideran como polimorfismos neutros que pueden transmitirse a las generaciones siguientes sin consecuencias de ningún tipo

(más allá de su utilización potencial como marcadores genéticos como, por ejemplo, los SNP o *single nucleotide polymorphisms*). En el caso de los embriones, la aparición de nuevas mutaciones durante las fases tempranas del desarrollo puede implicar que el individuo esté constituido por una mezcla (de proporciones variables) de dos o más estirpes celulares, lo que se denomina un mosaico genético, término que hace referencia a la heterogeneidad de su constitución genética (Acuna-Hidalgo et al., 2015).

Las mutaciones no son siempre detectables de manera inmediata, por ejemplo si aparecen sobre genes que no forman parte del repertorio de genes que se expresan en un determinado tejido o en estadios de desarrollo concretos, así como si se trata de mutaciones recesivas en estado de heterocigosis. Evidentemente, si la mutación aparece en una línea celular que contribuye a la formación de las gónadas, y por consecuencia de los gametos, aquella mutación podrá ser transmitida a las generaciones siguientes y todas las células deberían ser heterocigotas para la nueva mutación en el individuo original. Por otro lado, en los individuos adultos pueden aparecer nuevas mutaciones en aquellos tejidos con un alto recambio celular (*i.e.*, alta actividad mitótica), tales como la médula ósea, el intestino, los bronquios y la piel (además de las gónadas donde la multiplicación celular se da por el proceso de la meiosis). En algunos casos, las mutaciones somáticas pueden provocar la transformación de las células en células tumorales, especialmente cuando afectan a oncogenes, genes supresores de tumores, y genes de reparación del ADN (Benavides y Guénet, 2003).

2.2.1. Clasificación de las mutaciones:

El sistema de clasificación de las mutaciones que se describe a continuación es el señalado por Benavides y Guénet (2003). Según el tipo de célula que se ve afectada por la mutación, estas se pueden clasificar en:

a) Mutaciones de células somáticas: Estas mutaciones afectan a las células somáticas y como consecuencia pueden aparecer individuos mosaico con dos líneas celulares diferentes con distinto genotipo. Una vez que una célula sufre una mutación, todas las células que derivan de ella por divisiones mitóticas heredarán la mutación (*i.e.*, herencia celular). Un individuo mosaico originado por una mutación somática posee un grupo de células con un genotipo diferente al resto. Cuanto antes se haya dado la mutación en el desarrollo del individuo mayor será la proporción de células con distinto genotipo. En el supuesto de que la mutación se hubiera dado justo después de la primera división del cigoto (en estado de dos células), la mitad de las células del individuo adulto tendrían un genotipo y la otra mitad otro distinto. Las mutaciones que afectan solamente a las células de la línea somática no se transmiten a la siguiente generación.

b) Mutaciones de células germinales: Afectan a las células productoras de gametos y son transmitidas a la siguiente generación, teniendo una mayor importancia desde el punto de vista evolutivo. De hecho, estas son las mutaciones de interés para las investigaciones desarrolladas en la presente tesis doctoral.

En segundo lugar, podemos clasificar las mutaciones según el nivel genómico afectado:

a) Mutaciones moleculares: Se producen por una alteración en la estructura del ADN; generalmente se ve afectado un gen y se produce un cambio en la secuencia de nucleótidos. Estos cambios pueden darse por sustitución de bases, por inserción o deleción (es decir, cuando una base es añadida o eliminada y da origen a un cambio de pauta en la lectura del codón), o por transposición de bases.

b) Mutaciones cromosómicas: Son cambios o alteraciones que afectan la

estructura del cromosoma a nivel de algún segmento (generalmente se ve afectado más de un gen). Este tipo de mutaciones se puede presentar también por cambios en la disposición de los genes, pudiendo ser de varios incluyendo (1) por deleción (cuando se pierde un fragmento del cromosoma), (2) por duplicación (cuando se duplica un fragmento del cromosoma), (3) por adición (cuando se incorpora al cromosoma un grupo de nucleótidos), (4) por translocación (cuando un fragmento del cromosoma se une a otro cromosoma), y (5) por inversión (cuando un fragmento del cromosoma se invierte y no es leído correctamente).

c) Mutaciones genómicas: En este tipo de mutaciones, se ve afectada la dotación cromosómica del individuo, alterándose el número total de cromosomas con relación a la especie a la cual pertenece. Generalmente se produce por una separación anormal de los cromosomas durante la meiosis y pueden ser de dos tipos a su vez, (1) euploidía (cuando se ve afectado el juego completo de cromosomas), y (2) aneuploidía (cuando se afecta el número de cromosomas pero no la serie completa). Un ejemplo típico para las aneuploidías serían las trisomías comúnmente presentes en la especie humana, donde una de las más comunes es el síndrome de Down o trisomía 21.

También podemos clasificar las mutaciones según el cambio que ocasionan, pudiendo ser (1) perjudiciales para el individuo que las porta (muchas veces provoca la no viabilidad del cigoto y otras veces suelen originar enfermedades o malformaciones; *i.e.*, mutaciones letales, subletales y patógenas), (2) neutras (no proporcionan ningún perjuicio ni beneficio al individuo, aunque esto puede verse modificado si se originan cambios en el ambiente para los que la mutación pueda conllevar o no ventajas adaptativas), o (3) favorables (los individuos portadores de la mutación poseen ventajas adaptativas respecto a sus congéneres; el alelo mutante, con el tiempo y gracias a la selección natural, es posible que sustituya al alelo

original en la mayoría de los individuos que componen la población).

Finalmente, las mutaciones pueden diferenciarse según la forma como aparecen. Cuando una mutación se produce como consecuencia de la exposición a agentes mutagénicos químicos o físicos se denominan “mutaciones inducidas”, mientras que cuando se trata de una mutación que se produce de forma natural entonces es una “mutación espontánea”.

La información genética se transmite de padres a hijos, pero cada individuo puede tener a lo largo del genoma un importante número de mutaciones nuevas, que no se encontraban en la generación anterior. Estas mutaciones, a menudo llamadas *de novo*, contribuyen a la variabilidad genética y, también a largo plazo, a la evolución de la especie.

2.3. Recurrencia de las mutaciones

Como se ha comentado anteriormente, los distintos mecanismos implicados en la replicación, transcripción y traducción del ADN, así como en la división celular pueden cometer errores con cierta frecuencia. Cuando hablamos de poblaciones, el efecto de la mutación difiere de acuerdo con el interés puesto en el evento mutacional, tan raro que virtualmente sea único, o en un evento mutacional que recurra repetidamente. En términos generales la diferencia radica en la frecuencia con la que aparecerá un patrón mutante determinado y, en consecuencia, con la probabilidad que tendrá el mismo de transmitirse a la siguiente generación y alcanzar una frecuencia suficiente como para incidir a nivel poblacional.

a) Mutaciones no recurrentes: Si se considera un evento mutacional que da lugar a un solo gen mutado o a un cromosoma mutado en una población, su impacto potencial debería ser de poca importancia en la población a corto

plazo ya que el producto de la mutación única tiene una oportunidad infinitamente pequeña de sobrevivir en una población grande. La excepción se produce en el caso de que la mutación conlleve una ventaja selectiva (tanto si la selección es natural como artificial). De hecho, ya Fisher (1930) estableció que las mutaciones no recurrentes tenían una probabilidad baja de permanecer en las poblaciones. Esta dependía, en términos generales, del número de descendientes que contribuía el individuo mutante (*i.e.*, a mayor número de descendientes, mayor probabilidad de permanecer en la población).

b) Mutaciones recurrentes: Este tipo de mutaciones representa un fenómeno de interés particular para distintos ámbitos de la genética ya que se considera como un mecanismo biológico capaz de generar cambios en la frecuencia génica poblacional, independientemente de que la mutación implique o no algún tipo de ventaja selectiva. Cada evento mutacional se da regularmente y de manera repetida en distintos individuos de la población, con una frecuencia característica, impidiendo que pueda perderse por completo en poblaciones de tamaño moderado donde la deriva génica tendría un efecto muy secundario (Fisher, 1930; Lessa 2004). Este tipo de mutación provocaría cambios, generalmente lentos y difícilmente detectables a través de experimentos, en la frecuencia génica.

2.4. Influencia de la edad y el sexo de los progenitores sobre las mutaciones

Antes de profundizar en los vínculos entre la edad y la incidencia de nuevas mutaciones, debemos ser conscientes de las diferencias fisiológicas principales a nivel de la generación de cada línea germinal, dando lugar a los gametos en cada sexo. En el caso de los machos, la espermatogénesis se fundamenta en la multiplicación continua de las células basales una vez alcanzada la pubertad. Esta

multiplicación genera, por una parte, espermatogonias que sufren las sucesivas mitosis y las dos divisiones meióticas para producir los gametos, y por otra, células basales que persisten como base del ciclo, también sometidas regularmente a mitosis. Como consecuencia, las espermatogonias producidas a edades más tardías proceden de células basales que han sufrido un mayor número de divisiones mitóticas previas. Dado que cada división ofrece una cierta probabilidad de generar una nueva mutación puntual, resultaría plausible asumir un incremento en la acumulación de mutaciones puntuales con la edad en el caso del sexo masculino (Crow, 2000).

En el caso de las hembras se han destacado diferencias importantes. El proceso de la oogénesis inicia la meiosis durante el desarrollo fetal, pero los oocitos primarios quedan parados en la profase I ya antes del nacimiento. No será hasta que la hembra alcance la pubertad que uno o varios oocitos retomaran los procesos de la meiosis hasta finalizarla en cada ciclo reproductivo, sin requerir de divisiones adicionales en estas células u otras de precursoras (Murray et al., 2012). Dentro de este contexto, la probabilidad de acumular mutaciones se sugiere como claramente inferior que en el caso de los espermatozoides, al implicar un número mucho menor de mitosis. No obstante, también se ha sugerido que las mutaciones *de novo* pueden surgir con posterioridad a la gametogénesis y la fecundación, aunque se desconoce en gran medida la contribución relativa de cada uno de estos estadios reproductivos. Únicamente Acuna-Hidalgo et al. (2015) reportó una primera aproximación a este fenómeno, sugiriendo que una de cada quince mutaciones *de novo* no ocurrirían durante la formación del espermatozoide o el óvulo.

Múltiples autores han reportado que la función reproductiva del macho se ve claramente afectada por la edad (Callaway, 2012; Goriely y Wilkie, 2012). El envejecimiento reproductivo del macho es un fenómeno complejo y multifactorial, que implica no solo una disminución de la actividad sexual por alteraciones en la síntesis de hormonas esteroideas, sino también una disminución de la fertilidad por (1) un aumento del estrés oxidativo, (2) un incremento de la fragmentación del ADN,

y finalmente (3) un aumento considerable de mutaciones genómicas y cambios epigenéticos importantes (Vagnini et al., 2007; Cocuzza et al., 2008). Se ha sugerido que la tasa de mutaciones de la línea germinal en los varones humanos, especialmente en los hombres mayores, puede ser sustancialmente mayor que en las mujeres, principalmente porque en los hombres se dan muchas más divisiones de células germinales (Crow, 2000). Sin embargo, hay algunas excepciones y muchas variaciones ya que los distintos tipos de mutaciones siguen reglas diferentes. A partir del análisis evolutivo de la secuencia del genoma humano se infiere que la tasa global de mutación deletérea podría llegar a ser lo suficientemente alta como para influir de manera significativa en el bienestar humano. De hecho, algunos estudios sobre el impacto de la depresión endogámica en nuestra especie así lo corroborarían (Ober et al., 1999; Ceballos y Alvarez, 2013; Fareed y Afzal 2014), aunque existen mecanismos por los que el impacto de las mutaciones deletéreas podría ser mitigado (Crow, 2000; Fina, 2013).

De manera más específica, los aspectos reproductivos asociados con la edad avanzada en hombres se centran en un empeoramiento de los parámetros seminales clásicos (*i.e.*, disminución del volumen del eyaculado y motilidad espermática y, en menor grado, alteraciones de la morfología). Esto podría ser explicado por varios mecanismos asociados con el envejecimiento, deteriorando la calidad seminal e implicando deficiencias en la producción de testosterona (Johnson et al., 2015). De hecho, se han reportado cambios relevantes a nivel de los testículos y glándulas accesorias como (1) la reducción del volumen testicular (30%) y el número de células de Sertoli, (2) la disminución en el número de células de Leyding y de su respuesta secretora a la hormona luteinizante, (3) el incremento en número de células mesenquimales y tejido conectivo, (4) la esclerosis de los túbulos seminales, (5) la obstrucción vascular testicular (obstaculizando la llegada de gonadotropinas), y (6) los cambios en la glándula prostática y epidídimo (Wiecher y Cuéllar, 2007). A nivel del eje hipotálamo-hipofisiario, se puede observar también una disfunción neuroreguladora de la hormona liberadora de gonadotropina, así como una

disminución tanto de la amplitud como de la frecuencia de los pulsos secretores de hormona luteinizante (Wiecher y Cuéllar, 2007).

A todo lo citado anteriormente, debe considerarse también una correlación positiva entre la edad avanzada del padre y el daño en el ADN espermático (*i.e.*, fragmentación del ADN), detectable a partir de los 40 años (Alshahrani et al., 2014). Probablemente se debe a un mayor estrés oxidativo en el testículo del hombre de edad avanzada (*i.e.*, menor capacidad antioxidante), alteraciones en la apoptosis, defectos en el empaquetamiento del ADN, y alteraciones en la actividad de la telomerasa (Alshahrani et al., 2014).

Para el caso de las aneuploidías cromosómicas, la mayoría se originan *de novo*, por alteraciones en la meiosis de alguno de los progenitores, o errores mitóticos post-fecundación. Típicamente, los embriones resultantes no se implantarán o causarán abortos precoces, pero en un ~1% de los casos la gestación resultará viable. En el varón sano, se estima que el porcentaje de espermatozoides aneuploides es del 10%, pero se incrementa con la edad, principalmente a partir de 50 años. En la especie humana el riesgo se incrementa principalmente para la trisomía 21, doblándose para los nacidos de padres con más de 50 años (Nazer et al., 2008).

Centrándonos finalmente en las mutaciones, la causa principal del incremento de las mismas es la edad paterna avanzada según Callaway (2012) y Goriely y Wilkie (2012). De hecho, Kong et al. (2012) reportó que la tasa de mutaciones paternas aumentaba en un 4,3% al año, lo que implica duplicar la tasa de mutación cada 16 años y aumentarla ocho veces en 50 años. La espermatogénesis conlleva múltiples divisiones asimétricas premeióticas y se estima que de los 20 a los 50 años de edad, los espermatozoides pasan de sufrir 150 divisiones a 840, todo ello en un ambiente testicular cada vez más propenso al estrés oxidativo. Esto incrementa el riesgo de errores en la replicación y, por consiguiente, mutaciones. Se estima que las

mutaciones acumuladas en la línea germinal aumentan en dos cada año y el tipo más frecuentemente asociado a los trastornos sería la sustituciones de bases (Vagnini et al., 2007; Cocuzza et al., 2008).

Con la edad observamos también un incremento de alteraciones en los procesos epigenéticos fundamentales (*i.e.*, metilación del ADN, modificaciones de las histonas, o remodelación de la cromatina), pudiéndose desencadenar de forma inadecuada el mecanismo de “silenciado génico”, inactivando este la expresión de algunos genes. También estos mecanismos son la base de la impronta genómica en la que los genes se expresan de forma monoalélica (uno de los alelos, procedente de uno de los dos progenitores, está silenciado), perdiéndose la protección que proporciona el diploidismo frente a mutaciones recesivas (Vagnini et al., 2007; Cocuzza et al., 2008; Nazer et al., 2008; Johnson et al., 2015). Además se ha reportado una fuerte correlación entre la edad paterna avanzada la longitud de los telómeros de los cromosomas en la descendencia, lo que debería incrementar el riesgo de padecer algunos tipos de cáncer como el de mama o el glioblastoma (Maher et al., 2016).

En conjunto, la edad paterna avanzada se ha vinculado con una extensa lista de trastornos y patologías asociadas. Sin ánimo de ser exhaustiva, se podrían clasificar de la siguiente manera (McIntosh et al., 1995; Yang et al., 2007; Hultman et al., 2011):

a) Trastornos neurocognitivos: Esquizofrenia, trastorno bipolar y autismo. La relación de estos trastornos con la edad avanzada de los padres, aunque no de forma unánime, si es de las aceptadas con mayor consenso. Se han asociado a la acumulación de nuevas mutaciones, pero parecen también jugar un papel destacado los fenómenos epigenéticos, en especial la impronta. La esquizofrenia es uno de los que mayor influencia genética tiene, hasta el punto de atribuirse a la avanzada edad de los padres en un 25% de los casos, y

un aumento del riesgo de 1,47 de padecerla con cada incremento de 10 años en la edad paterna. También el trastorno bipolar se ha relacionado aunque en menor medida, mientras que se estima que el riesgo de autismo se dobla para varones de más de 50 años en relación con los de más de 30 (Hultman et al., 2011).

b) Procesos de herencia autosómica dominante (aunque generalmente se presentan como consecuencia de mutaciones de novo). Destaca por su fuerte asociación con la edad paterna avanzada la acondroplasia. No obstante, existen otros como el síndrome de Apert, el síndrome de Marfan, la neurofibromatosis I, la osteogénesis imperfecta y el retinoblastoma (McIntosh et al., 1995).

c) Malformaciones congénitas: El efecto de edad paterna avanzada sobre una mayor incidencia de defectos congénitos resulta muy incierto. Los grandes estudios epidemiológicos son escasos y sin uniformidad en sus observaciones. Así, mientras que unos no encuentran asociación, otros sí, pero difieren en el tipo de defecto y, en todo caso, la señalan como discreta. Destacan los defectos del tubo neural, cardíacos, atresia traqueo-esofágica, defectos musculo-esqueléticos, y síndrome de Down (McIntosh et al., 1995; Yang et al., 2007).

d) Neoplasias: Se ha documentado un incremento del riesgo en la descendencia para algunos tipos neoplásicos como la leucemia, los tumores del sistema nervioso central, y los tumores de mama (McIntosh et al., 1995; Maule et al., 2007).

A pesar que algunas investigaciones recientes han avanzado en este campo para el caso de los humanos (Hultman et al., 2011; Kong et al., 2012; Keightley, 2012), no existe aún un criterio universal para considerar a partir de qué edad debería considerarse que existe un riesgo asociado excesivo en el caso de los humanos, y

mucho menos aún para los animales domésticos o de laboratorio. De cualquier forma, son muchas y muy contundentes las evidencias mostradas para considerar la edad como un problema importante que actúa sobre la fertilidad, así como los riesgos mutacionales asociados que se afrontan en caso de edad paterna avanzada, sobretodo dentro del contexto de la especie humana (Hultman et al., 2011; Callaway, 2012; Goriely y Wilkie, 2012; Kong et al., 2012).

3. LAS MUTACIONES EN EL CONTEXTO DE LA GENÉTICA

Hasta este punto se han tratado las mutaciones como eventos genéticos individuales, considerando que aparecen como un fenómeno aislado, con o sin repercusiones en el organismo que la presenta; sin embargo, las mutaciones espontáneas pueden implicar connotaciones importantes sobre los cambios que acontecen en las poblaciones, y pueden ser consideradas cuando se pretende dilucidar la base genética de la evolución con fines experimentales, o bien para el mejoramiento genético de las características de interés ganadero. En este mismo orden de ideas, podemos inferir que la mutación es la fuente última de variación genética y, por lo tanto, de la evolución de las especies (Hill 1982a,b; Keightley, 1998). En esta sección, se discutirán algunos aspectos aplicados de las mutaciones desde un punto de vista cuantitativo, pero fundamentalmente, sobre el modo en como se incorporan dichas mutaciones a los modelos analíticos específicos del campo de la genética.

El proceso de mutación (*i.e.*, cualquier cambio que se produce en el genoma con potencial de transmitirse a las nuevas generaciones), aunque muy lentamente, va generando nuevos alelos que entran a formar parte del acervo génico de la población (Caballero, 2014). Estos nuevos alelos incidirán temporal o permanentemente sobre la variabilidad genética, siendo este último parámetro la materia prima para el estudio de la genética de poblaciones. De hecho y desde un punto de vista puramente teórico, una población uniforme en su composición genética (*i.e.*, no variable) no muestra cambios relevantes a largo plazo, con lo cual no puede evolucionar. En la práctica, las nuevas mutaciones aparecen para aportar variación, se puede inferir que las nuevas mutaciones serán relativamente numerosas, pero están repartidas en un genoma extremadamente grande. ¿Cómo se puede determinar la magnitud de su contribución? Para responder a esta pregunta se deben integrar estas mutaciones dentro de los modelos analíticos específicos del campo de la genética.

3.1. Breve reseña sobre la evolución de la selección animal

El mejoramiento genético de las poblaciones se centra en la implementación de iniciativas basadas en distinguir los individuos genéticamente superiores para los caracteres de interés, siendo estos los progenitores de la siguiente generación. No obstante, resulta importante destacar que el mérito genético no puede ser observado directamente en los animales, ya que el fenotipo final recibe otras contribuciones ambientales que enmarcaran el impacto directo de los genes; la base genética debe inferirse a partir de las observaciones hechas en los candidatos a selección, o en sus parientes. Esto presenta al menos tres problemas estadísticos, (1) determinar si los caracteres de interés tienen base genética (*i.e.*, estimación de los parámetros genéticos de varianza; Groeneveld et al., 1990), (2) obtener métodos razonablemente precisos para inferir el mérito genético (*i.e.*, desarrollo de modelos de evaluación genética precisos; Henderson, 1973), y (3) decidir qué hacer con los animales que tengan mejor mérito genético (e.g., diseño de planes de selección óptimos).

Se puede asumir que las primeras iniciativas ejecutadas por las poblaciones humanas para adaptar las poblaciones de animales domésticos a las necesidades de sus cuidadores se produjeron durante el periodo neolítico, hace miles de años, coincidiendo con los eventos de domesticación (Giuffra et al., 2000; Chessa et al., 2009; Scheu et al., 2015). Este proceso de domesticación involucró cambios genéticos de suma importancia, en los que las especies salvajes que existían en la naturaleza fueron, por acción del hombre, diferenciadas en poblaciones domésticas. Esta diferenciación se trató de un proceso lento y progresivo, pudiendo evolucionar de tal manera que, en muchos casos, se llegó a una separación reproductiva completa. En esos momentos, cualquier tipo de manipulación genética se ejecutó de forma mayoritariamente inconsciente. Los cambios obtenidos como resultados de tales maniobras fueron conducidos de una forma totalmente empírica; no obstante,

varios investigadores sugirieron que los cambios logrados fueron apoyados fundamentalmente en la observación del parecido entre parientes. Este hecho condujo a la elección como reproductores de aquellos animales y plantas que parecían ofertar un mejor ajuste a los requerimientos de la humanidad (i.e., mayor producción y calidad de los productos, potencia de trabajo, o docilidad de los animales; Diamond, 1999). No obstante, no es hasta el Libro del Génesis en el que podemos encontrar las primeras referencias escritas sobre la selección empírica de los animales domésticos. También otros textos no religiosos de civilizaciones como la Romana o la Cartaginesa hacen alguna referencia a estos procedimientos.

No fue hasta muchos siglos después, con los resultados obtenidos por Mendel, que se empezaron a sentar las bases de la genética moderna (Mendel, 1866). Paralelamente, resulta curioso mencionar que Charles R. Darwin, coetáneo de Mendel, desconocía sus descubrimientos cuando desarrollo toda su teoría de la evolución de las especies. Aun así, fue en esos trabajos donde empezó a usarse el término “selección”, siempre dentro del contexto de la evolución de las poblaciones (Darwin, 1859). Darwin se dio cuenta que muchas especies de animales y plantas domesticadas destacaban por características concretas que se habían desarrollado mediante selección artificial. Ese detalle resultó el punto de partida para una plétora de programas de selección que han dado lugar a miles de estirpes, razas, líneas y otras entidades genéticas.

El siguiente avance destacable se produjo a principios del siglo XX, cuando Fisher (1918a) desarrolló la teoría necesaria para el análisis de la varianza (ANOVA), representando este el punto de partida para un análisis eficiente de los efectos aleatorios dentro del contexto de los modelos mixtos, sobretudo para la contribución del gómonente genético (Fisher, 1918b). Posteriormente se desarrollaron los índices de selección para el mejoramiento genético de plantas (Smith, 1936) a partir del concepto de la función discriminante (Fisher, 1936). Y poco después, Hazel (1943) adaptó esos procedimientos para la evaluación genética de animales, generalizándose

como el método de elección en la mayoría de programas de selección. Llegados a este punto resulta importante destacar también las contribuciones de otros investigadores como Lush (1940).

El siguiente avance, la evaluación mediante modelos BLUP (del inglés *best linear unbiased prediction*) se desarrolló poco después, en 1950, aunque no se divulgó a nivel internacional hasta dos décadas más tarde (Henderson, 1973). Esta metodología permitía predecir el valor genético de los animales a partir de los registros fenotípicos y la genealogía, incluyendo y corrigiendo también por las demás causas ambientales conocidas de variación. Resulta importante destacar que las ventajas del BLUP se deben fundamentalmente a la inclusión de la matriz de parentescos (Wright, 1922) en forma de inversa (Henderson, 1976; Quaas, 1976). Esta matriz incluye todas las relaciones conocidas de parentesco. A partir de aquí, la parametrización estándar del BLUP se modificó para incluir efectos de dominancia (Henderson, 1985; Hoeschele y VanRaden, 1991), epistasia (de los Campos et al., 2009; Casellas et al., 2014), consanguinidad (Gulisija et al., 2006; Casellas et al., 2008), e incluso nuevas mutaciones (Wray, 1990; Casellas y Medrano, 2008).

En términos generales, estos modelos de evaluación genética mediante ecuaciones BLUP, són estructuras matemáticas dotadas de los siguientes componentes (Falconer y Mackay, 1996):

- a)** Una función matemática que relaciona las observaciones fenotípicas con los efectos fijos y aleatorios del modelo. En términos generales, se acostumbran a asumir modelos lineales (Henderson, 1973), aunque también se dan excepciones importantes como los modelos de riesgos proporcionales (Ducrocq et al., 1988; Casellas, 2007) o los modelos umbral (Sorensen et al., 1995). Los efectos aleatorios pueden incluir componentes genéticos (*e.g.*, valores genéticos aditivos, de dominancia, o desviaciones epistáticas), así como efectos ambientales permanentes como efectos maternos ambientales

(Quintanilla et al., 1999). Todos estos factores contribuyen a la covarianza entre parientes.

b) Parámetros de dispersión genética y ambiental, tales como componentes de varianza y covarianza. Estos últimos aparecen en modelos multivariados, o en aquellas situaciones en las cuales debe incorporarse una estructura multivariada a un modelo para una respuesta univariada. Sería el caso de modelos con efectos genéticos directos y maternos, los cuales pueden estar vinculados mediante una covarianza genética (Quintanilla et al., 1999).

c) Supuestos sobre la forma de la distribución conjunta de las observaciones y de los efectos aleatorios (en un contexto Bayesiano, los supuestos se aplican a la distribución conjunta de todos los parámetros desconocidos y los datos). En este caso, la asunción más típica es la de normalidad de los datos fenotípicos, aunque nuevamente, se han propuesto variaciones sustanciales para fenotipos específicos (Ducrocq et al., 1988; Casellas, 2007).

El desarrollo paralelo que sufrieron las tecnologías laboratoriales para el genotipado de genes y marcadores genéticos permitieron adaptar los programas de mejora genética con el objetivo de integrar la selección directa sobre unos pocos genes mayores (Dekkers, 2004). Este tipo de selección se conoció, en términos generales, como selección asistida por marcadores (del inglés *marker assisted selection* (MAS)) y, aunque en algunos casos se implementó durante periodos largos de tiempo (Eikelenboom y Minkema, 1974), su impacto real en las poblaciones domesticas de ganado fue realmente limitada. De hecho, se han reportado infinidad de *quantitative trait loci* (Andersson, 2001), y muy pocos de ellos han llegado a ser implementado de manera efectiva como criterios de selección en los programas reales de mejora genética. Quizá una excepción destacable sería el gen del receptor de estrógenos (Rothschild y Soller, 1997), que se ha usado en múltiples líneas porcinas comerciales para incrementar la prolificidad (Dekkers, 2004). En cualquier

caso, BLUP y MAS han sido las herramientas principales para la mejora genética animal durante las últimas décadas, al menos hasta el disparo de salida para los procedimientos de selección genómica (Meuwissen et al., 2001).

Basándose en la disponibilidad (o en la futura disponibilidad) de miles de marcadores de tipo *single nucleotide polymorphism* (SNP) en la mayoría de especies domésticas (e incluso en la humana), Meuwissen et al. (2001) postularon que la varianza genética aditiva originada a partir de algunas docenas o unos pocos centenares de mutaciones en el genoma (Hayes y Goddard, 2001) se podía capturar a partir de un panel denso de SNP, algunos de estos marcadores estando parcial o totalmente en desequilibrio de ligamiento con las mutaciones causales. Eso se implementó rápidamente mediante modelos estadísticos complejos, incidiendo inicialmente en los programas de selección de vacuno lechero (Lund et al., 2011; Wiggans et al., 2011) y porcino (Forni et al., 2011; Ostersen et al., 2011). Dentro de ese contexto, se evidenció la utilidad de incluir la información procedente de genotipados masivos en los procedimientos de evaluación genética. Por ejemplo, para los sementales en testaje, la precisión de sus valores mejorantes para producción de leche se incrementó de ~27% al ~54% (VanRaden et al., 2009). Evidentemente, en la actualidad se dispone de múltiples aproximaciones distintas para abordar este tipo de evaluaciones genómicas (Gianola, 2013). Típicamente, estas se diferencian por el abordaje general de la información disponible, requiriendo primero la estimación de los efectos aditivos asociados a cada SNP para posteriormente sustituir esas estimas en los genotipos de los nuevos animales de la población (i.e., aproximaciones tipo *two-step*; Gianola et al., 2009; Habier et al., 2011), o resolviendo todo el sistema en un único paso (i.e., aproximaciones tipo *one-step*; Aguilar et al., 2010) gracias a que la información genómica se tiende a incluir en la matriz de parentescos (VanRaden, 2008). Dentro de este segundo grupo encontramos el BLUP genómico (Gianola et al., 2003), el BLUP genómico mutacional (Casellas et al., 2013), el LASSO Bayesiano (Yi y Xu, 2008; de los Campos et al., 2009), la regresión por espacios de Hilbert (Gianola y van Kaam, 2008), o el *boosting* (González-Recio et al., 2013).

Además, se pueden adaptar para permitir la evaluación conjunta de animales tanto genotipados como no genotipados (Legarra et al., 2009; Misztal et al., 2009; Aguilar et al., 2010).

3.2. La variabilidad mutacional *de novo* en las poblaciones

El impacto de las nuevas mutaciones puede ser descrito desde diversos puntos de vista, incluso desde el ámbito específico de la genética, y tienen consecuencias evolutivas intrínsecas basadas en la variabilidad atribuida a las especies (Behrman y Kirkpatrick, 2011; Nuzul et al., 2014). No obstante, las mutaciones son capaces de lo mejor y lo peor desde un punto de vista biológico, originando organismos mejor adaptados a los cambios ambientales (Shen et al., en prensa) a la vez que amenazando la supervivencia a través de efectos deletéreos (Lynch et al., 1995). Las mutaciones espontáneas, originadas de manera continua, aportan nuevos alelos que contribuyen a la variación genética y así, permiten una respuesta sostenida a la selección, tanto en animales (Caballero et al., 1991; Keightley, 1998) como en plantas (Hill, 2007).

El estudio genético de las nuevas mutaciones desde un punto de vista cuantitativo gira de manera fundamental en torno a la variabilidad genética que originan, ya que es precisamente en términos de varianza como se acostumbran a descomponer las distintas contribuciones tanto genéticas como ambientales sobre las poblaciones y los fenotipos analizados (Falconer, 1996). Las primeras evidencias de nuevas mutaciones con efectos grandes en líneas experimentales de animales de laboratorio se reportaron durante la segunda mitad del siglo XX (Castle, 1941; Macarthur, 1949; Yoo, 1980; Bradford y Faula, 1984), aunque también disponemos de fenómenos parecidos en las especies ganaderas (Smith and Bampton 1977; Jackson and Green 1993). No obstante, este tipo de mutaciones debería verse como una excepción, ya que en la mayoría de casos su contribución individual sobre la

variabilidad genética de las poblaciones es mínima, aunque relevante desde un punto de vista evolutivo (Hill, 1982a,b; Morgan et al. 2014). Dentro de este contexto, se acostumbra a definir la varianza aditiva mutacional (σ_m^2), entendiéndola como la contribución mutacional de nueva variabilidad genética aditiva por generación (Hill, 1982a). Además, esto permite también definir una heredabilidad mutacional de acuerdo con la siguiente formulación,

$$h_m^2 = \sigma_m^2 / \sigma_p^2$$

donde σ_p^2 no es más que la varianza fenotípica total del carácter analizado. De acuerdo con los trabajos de revisión publicados por Lynch (1988) y Houle et al. (1996), la magnitud esperada de h_m^2 se sitúa entre 0,0001 y 0,05 tanto en animales como en cultivos vegetales. No obstante, en la mayoría de casos las estimas se obtuvieron a partir de diseños experimentales centrados en la tasa de divergencia entre sublíneas de una misma población consanguínea (Festing, 1973; Mackay et al., 1994). Esos estudios tomaban algunas asunciones francamente restrictivas en cuanto al equilibrio entre mutación y cambio genético, lo que no necesariamente debería cumplirse en las poblaciones experimentales (Hill, 1982a). Además, se requiere de periodos muy largos de tiempo para la generación de las sublíneas, mientras que el análisis de la respuesta a la selección acostumbra a ignorar la información relativa a las covarianzas entre parientes de una misma sublínea, una parte de las cuales puede ser puramente genética (Keightley y Hill, 1992). Esto se constató en los trabajos de Keightley y Hill (1992), Keightley (1998), Casellas y Medrano (2008) y Casellas et al. (2010), desarrollados a partir de la parametrización propuesta por Wray (1990) sobre la matriz de parentescos mutacionales aditivos. En cualquier caso, los resultados obtenidos no diferían de manera significativa en relación con los revisados por Lynch (1988) y Houle et al. (1996), aunque se situaban en su límite superior. El planteamiento de Wray (1990) asumía un escenario específico donde las nuevas mutaciones, con efectos aditivos pequeños sobre el fenotipo, aparecían aleatoriamente en todas las generaciones. Estas nuevas mutaciones eran heredadas

por los descendientes siguiendo un patrón aditivos infinitesimal y, como consecuencia, se podían modelar fácilmente mediante ciertas adaptaciones en la matriz de parentesco (Wright, 1922). Esta parametrización proporcionaba una herramienta muy útil para dilucidar los aspectos más relevantes inherentes a las nuevas mutaciones, sus patrones de emergencia, y su contribución sobre la varianza genética total de la población. A partir de las modificaciones propuestas por Casellas y Medrano (2008), se facilitó el análisis de este fenómeno biológico desde un punto de vista Bayesiano, extendiéndolo incluso para capturar efectos mutacionales dentro de un contexto de BLUP genómico con información de genotipado de miles de marcadores genéticos (Casellas et al., 2013). Esta ampliación se llevó a cabo una vez evidenciadas las limitaciones de los modelos BLUP genómico estándares ante nuevas mutaciones aditivas en la población, incidiendo directamente sobre la variabilidad del carácter (Casellas y Varona, 2011).

Además, resulta lógico asumir que parte de las nuevas mutaciones pueden no únicamente generar varianza genética aditiva por si mismas, sino también al interaccionar con otros polimorfismos existentes en el genoma. Este fenómeno se conoce como epistasia y ha sido muy poco estudiado debido tanto (1) a las complejidades metodológicas que comporta (Crow y Kimura, 1970; Goodnight, 1987, 1988; Wade, 1992; Cheverud & Routman, 1995), como (2) a la poca contribución que se le asume sobre la covarianza genética entre individuos emparentados (Cockerham, 1954; Hayman y Mather, 1955; Falconer y Mackay, 1996). Los resultados experimentales han arrojado evidencias en muchos casos controvertidas e incluso antagónicas (Simons y Crow, 1977; Barker, 1979), aunque en algunos casos se han evidenciado contribuciones importantes sobre la varianza fenotípica total (Peripato et al., 2005; Leamy et al., 2008). Para epistasia generada a partir de nuevas mutaciones, quizá el único resultado disponible sea el reportado por Casellas et al. (2014), llegando a explicar alrededor del 10% de la variabilidad total para el peso a las nueve semanas de edad en dos estirpes distintas de ratones de laboratorio.

En cualquier caso, las estimas reportadas para la varianza mutacional asumen un comportamiento homogéneo en todas las generaciones, e incluso a lo largo de la vida de los reproductores. Eso puede no ser cierto tal como se ha mencionado con anterioridad; no obstante, la probable evolución de la varianza mutacional a medida que envejecen los individuos es un fenómeno escasamente analizado desde el punto de vista de la genética cuantitativa. Algunos estudios recientes en la especie humana han sugerido vínculos importantes entre la edad paterna y la tasa de mutación (Goriely and Wilkie, 2012; Forster et al., 2015), exponiendo a los descendientes a una mayor incidencia de enfermedades como autismo, esquizofrenia o ciertos tipos de carcinomas (Bertram, 2000; Callaway, 2012; Kong et al., 2012).

BIBLIOGRAFÍA

Acuna-Hidalgo R, Bo T, Kwint MP, van de Vorst M, Pinelli M, Veltman JA, Hoischen A, Vissers LELM, Gilissen C. 2015 Post-zygotic point mutations are an underrecognized source of de novo genomic variation. *Am J Hum Genet* **97**: 67-74.

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. 2010. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci* **93**: 743-752.

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. 2002. *Molecular Biology of the Cell*. Garland Science, New York, NY, EUA.

Alshahrani S, Agarwal A, Assidi M, Abuzeadah AM, Durairajanayagam D, Ayaz A, Sharma R, Sabanegh E. 2014. Infertile men older than 40 years are at higher risk of sperm DNA damage. *Reprod Biol Endocrinol* **12**: 103.

Andersson L. 2001. Genetic dissection of phenotypic diversity in farm animals. *Nat Rev Genet* **2**: 130-138.

Avery OT, MacLeod CM, McCarty M. 1944. Studies on the chemical nature of the substance inducing transformation on Pneumococcal types. Induction of transformation by a desocytiribonucleic acid fraction isolated from Pneumococcus type III. *J Exp Med* **79**: 137-158.

Barker JS. 1979. Interlocus interactions: a review of experimental evidence. *Theor Popul Biol* **16**: 323-346.

Behrman KD, Kirkpatrick M. 2011. Species range expansion by beneficial mutations. *J Evol Biol* **24**: 665-675.

Benavides F, Guénet J-L. 2003. *Manual de genética de roedores de laboratorio*.

Principios básicos y aplicaciones. Universidad de Alcalá, Alcalá de Henares, España.

Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci USA* **74**: 3171-3175.

Bertram J. 2000. The molecular biology of cancer. *Mol Aspects Med* **21**: 167-223.

Bradford GE, Famula TR. 1984. Evidence for a major gene for rapid postweaning growth in mice. *Genet Res* **44**: 293-298.

Caballero A. 2014. Conservation genetics: applying evolutionary concepts to the conservation of biological diversity. *Sci Studies J Ann Rev* **4**: 73-77.

Caballero A, MA Toro, C López-Fanjul. 1991. The response to artificial selection from new mutations in *Drosophila melanogaster*. *Genetics* **127**: 89-102.

Callaway E. 2012. Genome study may explain links between paternal age and conditions such as autism. *Nature* **488**: 439.

Campbell N, Reece JB. 2011. *Biology*. Benjamin Cummings, Boston, MA, EUA.

Casellas J. 2007. Bayesian inference in a piecewise Weibull proportional hazards model with unknown change points. *J Anim Breed Genet* **124**: 176-184.

Casellas J, Caja G, Piedrafita J. 2010. Accounting for additive genetic mutations on litter size in Ripollesa sheep. *J Anim Sci* **88**: 1248-1255.

Casellas J, Esquivelzeta C, Legarra A. 2013. Accounting for new mutations in genomic prediction models. *J Dairy Sci* **96**: 5389-5402.

Casellas J, Gianola D, Medrano JF. 2014. Bayesian analysis of additive epistasis arising from new mutations in mice. *Genet Res* **96**: e008.

Casellas J, Medrano JF. 2008. Within-Generation Mutation Variance for Litter Size in Inbred Mice. *Genetics*. **179**: 2147-2155.

Casellas J, Varona L. 2011. Effect of mutation age on genomic predictions. *J Dairy Sci* **94**: 4224-4229.

Casellas J, Varona L, Ibáñez-Escriche N, Quintanilla R, Noguera JL. 2008. Skew distribution of founder-specific inbreeding depression effects on the Longevity of Landrace sows. *Genet Res* **90**: 499-508.

Castle WE. 1941. Influence of certain color mutations on body size in mice, rats, and rabbits. *Genetics* **26**: 177-191.

Ceballos FC, Alvarez. 2013. Royal dynasties as human inbreeding laboratories: the Habsburgs. *Heredity* **111**: 114-121.

Chargaff ED. 1952. On the deoxyribonucleic acid content of sea urchin gametes. *Experientia* **8**: 143-145.

Chargaff ED, Lipzhitz R, Green C. 1952. Composition of the deoxypentose nucleic acids of four genera of sea-urchin. *J Biol Chem* **195**: 155-160.

Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, Mainland I, Kao RR, Pemberton JM, Beraldi D, Stear M, Alberti A, Pittau A, Iannuzzi L, Banbazi MH, Kazwala R, Zhang Y-P, Arranz JJ, Ali BA, Wang Z, Uzun M, Dione M, Olsaker I, Holm L-E, Saarma U, Ahmad S, Marzanov N, Eythorsdottir E, Holland MJ, Ajmone-Marsan P, Bruford MW, Kantanen J, Spencer TE, Palmarini M. 2009. Revealing the history of sheep domestication using retrovirus integrations. *Science* **24**: 532-536.

Cheverud JM, Routman EJ. 1995. Epistasis and its contribution to genetic variance components. *Genetics* **139**: 1455-1461.

Chow LT, Gelinas RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**: 1-8.

Cockerham CC. 1954. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**: 859-882.

Cocuzza M, Athayde KS, Agarwal A, Sharma R, Pagani R, Lucon AM, Srougi M, Hallak J. 2008. El aumento relacionado con la edad de especies reactivas de oxígeno en semen puro en hombres sanos fértiles. *Urología* **71**: 490 - 494.

Coop G, Przeworski M. 2007. An evolutionary view of human recombination. *Nat Rev Genet* **8**: 23-34.

Crow JF. 2000. The origins, patterns and implications of human spontaneous mutations. *Nat Rev Genet* **1**: 40-47.

Crow JF, Kimura M. 1970. *An Introduction to Population Genetics Theory*. Harper and Row, New York, NY, EUA.

Dahm R. 2008. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum Genet* **122**: 565-581.

Darwin CR. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, Londres, Reino Unido.

Dekkers JCM. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J Anim Sci* **82**(E supl.): E313-E328.

de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Weigel K, Cotes JM. 2009. Predicting quantitative traits with regression models for dense

molecular markers and pedigree. *Genetics* **182**: 375-385.

Diamond J. 1999. *Guns, germs, and steel*. Norton Press, New York, NY, EUA.

Ducrocq V, Quaas RL, Pollak EJ. 1988. Length of productive life of dairy cows. 1. Justification of a Weibull model. *J Dairy Sci* **71**: 3061-3070.

Eikelenboom G, Minkema D. 1974. Prediction of pale, soft, exudative muscle with non-lethal test for the halothane-induced porcine malignant hyperthermia syndrome. *Tijdschr Diergeneeskde* **99**: 421-427.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet* **8**: 610-618.

Falconer DS, Mackay TFC. 1996. *Introduction to quantitative genetics*. Longman Group, Londres, Reino Unido.

Fareed M, Afzal M. 2014. Evidence of inbreeding depression on height, weight, and body mass index: A population-based child cohort study. *Am J Hum Biol* **26**: 784-795. *sis of subline divergence in the shape of the mandible in C57BL/Gr mice. Genet Res* **21**: 121-132.

Festing M. 1973. A multivariate analysis of subline divergence in the shape of the mandible in C57BL/Gr mice. *Genet Res* **21**: 121-132.

Fina M. 2013. *Optimització de l'avaluació genètica de la raça bovina Bruna dels Pirineus*. Tesis Doctoral, Universitat Autònoma de Barcelona, Bellaterra, Barcelona. España.

Fisher RA. 1918a. Studies in crop variation. I. An examination of the yield of dressed grain from Broadbalk. *J Agric Sci* **11**: 107-135.

Fisher RA. 1918b. The correlation between relatives on the supposition of Mendelian

inheritance. *Phil Trans Royal Soc Edinburgh* **52**: 399-433.

Fisher RA. 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford, Reino Unido.

Fisher RA. 1936. The use of multiple measurements in taxonomic problems. *Ann Eugenics* **7**: 179-188.

Forni S, Aguilar I, Misztal I. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet Sel Evol* **43**: 1.

Forster P, Hohoff C, Dunkelmann B, Schürenkamp M, Pfeiffer H, Neuhuber F, Brinkmann B. 2015. Elevated germline mutation rate in teenage fathers. *Proc R Soc B* **282**: 1-7.

Franklin R, Gosling RG. 1953. Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature* **171**: 156-157.

Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. 2004. A census of human cancer genes. *Nat Rev Genet* **4**: 177-183.

Gianola D. 2013. Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics* **194**: 573-596.

Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando RL. 2009. Additive genetic variability and the Bayesian alphabet. *Genetics* **183**: 347-363.

Gianola D, Pérez-Enciso M, Toro MA. 2003. On marker-assisted prediction of genetic value: beyond the ridge. *Genetics* **163**: 347-365.

Gianola D, van Kaam JBCHM. 2008. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* **178**: 2289-

2303.

Giuffra E, Kijas JMH, Amarger V, Carlborg Ö, Jeon J-T, Andersson L. 2000. The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* **154**: 1785-1791.

González-Recio O, Jiménez-Montero JA, Alenda R. 2013. The gradient boosting algorithm and random boosting for genome-assisted evaluation in large data sets. *J Dairy Sci* **96**: 614-624.

Goodnight CJ. 1987. On the effect of founder events on epistatic genetic variance. *Evolution* **41**: 80-91.

Goodnight CJ. 1988. Epistasis and the effect of founder events on the additive genetic variance. *Evolution* **42**: 399-403.

Goriely A, Wilkie AO. 2012. Paternal age effect mutations and selfish spermatogonial selection: causes and consequences for human disease. *Am J Hum Genet* **90**: 175–200.

Groeneveld E, Kovac M, Wang T. 1990. PEST, a general purpose BLUP package for multivariate prediction and estimation. Páginas 488-491 en *Proc 4th World Congr Genet Appl Anim Prod*, Edimburgo, Reino Unido.

Guenther CA, Tasic B, Luo L, Bedell MA, Kingsley DM. 2014. A molecular basis for classic blond hair color in Europeans. *Nat Genet* **46**: 748-752.

Gulisija D, Gianola D, Weigel KE, Toro MA. 2006. Between-founder heterogeneity in inbreeding depression for production in Jersey cows. *Livest Sci* **104**: 244-253.

Habier D, Fernando RL, Kizilkaya K, Garrick DJ. 2011. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* **12**: 186.

Hayes D, Goddard ME. 2001. The distribution of effects of genes affecting quantitative traits in livestock. *Genet Sel Evol* **33**: 209-229.

Hayman BI, Mather K. 1955. The description of genic interaction in continuous variation. *Biometrics* **10**: 69-82.

Hazel LN. 1943. Genetic basis for construction of selection indices. *Genetics* **28**: 476-390.

Henderson CR. 1973. Sire evaluations and genetic trends. Páginas 10-41 en *Proc Anim Breed Genet Symp in Honor of Jay L Lush*. American Society of Animal Science and American Dairy Science Association, Champaign, Illinois, EUA.

Henderson CR. 1976. A simple way for calculating the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* **32**: 69-83.

Henderson CR. 1985. Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. *J Anim Sci* **60**: 111-117.

Hill WG. 1982a. Rates of change in quantitative traits from fixation of new mutations. *Proc Natl Acad Sci USA* **79**: 142-145.

Hill, WG. 1982b. Predictions of response to artificial selection from new mutations. *Genet Res* **40**: 255-278.

Hill WG. 2007. A century of corn selection. *Science* **307**: 683–684.

Hoeschele I, VanRaden PM. 1991. Rapid inversion of dominance relationship matrices for noninbred populations by including sire by including dam subclass effects. *J Dairy Sci* **74**: 557-569.

Holliday R. 1964. A mechanism for gene conversion in fungi. *Genet Res* **5**: 282-304.

Houle D, Morikawa B, Lynch M. 1996. Comparing mutational variabilities. *Genetics* **143**: 1467-1483.

Hultman CM, Sandin S, Levine SZ, Lichtenstein P, Reichenberg A. 2011. Advancing paternal age and risk of autism: new evidence from a population-based study and a meta-analysis of epidemiological studies. *Mol Psychiatry* **16**: 1203-1212.

Jackson SP, Green RD. 1993. Muscle trait inheritance, growth performance and feed efficiency of sheep exhibiting a muscle hypertrophy phenotype. *J Anim Sci* **71**: 241 (Abstr.).

Johnson SL, Dunleavy J, Gemmell NJ, Nakagawa S. 2015. Consistent age-dependent declines in human semen quality: A systematic review and meta-analysis. *Ageing Res Rev* **19**: 22 -33.

Keightley PD. 1998. Genetic basis of response to 50 generations of selection on body weight in inbred mice. *Genetics* **148**: 1931-1939.

Keightley PD. 2012. Rates and Fitness Consequences of New Mutations in Humans. *Genetics*. **190**: 295-304.

Keightley PD, Hill WG. 1992. Quantitative genetic variation in body size of mice from new mutations. *Genetics* **131**: 693-700.

Kodym A, Afza R. 2003. Physical and chemical mutagenesis. *Meth Mol Biol* **236**: 189-204.

Kong A, Frigge M, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson S, Sigurdsson A, Jonasdottir A, Jonasdottir A, et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**: 429-550.

Kossel A, Steudel HZ. 1903. Weitere untersuchungen über das cytosin. *Physiol Chem*

38: 49-51.

Leamy LJ, Pomp D, Lightfoot JT. 2008. An epistatic genetic basis for physical activity traits in mice. *Genetics* **99**: 639-646.

Lee RC, Feinbaum RL, Amrbos V. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementary to *lin-14*. *Cell* **75**: 843-854.

Legarra A, Aguilar I, Misztal I. 2009. A relationship matrix including full pedigree and genomic information. *J Dairy Sci* **92**: 4656-4663.

Lessa E. 2004. *Guía de Estudio de Genética de Poblaciones*. Universidad Montevideo, Montevideo, Uruguay.

Lund MS, de Roos APW, de Vries AG, Druet T, Ducrocq V, Fritz S, Guillaume F, Gulbrandtsen B, Liu Z, Reents R, Schrooten C, Seefried F, Su G. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet Sel Evol* **43**: 43.

Lush JL. 1940 Intra-sire correlations or regressions of offspring on dam as a method of estimating heritability of characteristics. *Proc Am Soc Anim Prod* **1940**: 293-301.

Lynch M. 1988. The rate of polygenic mutation. *Genet Res* **51**: 137-148.

Lynch M, Conery J, Bürger R. 1995. Mutation accumulation and the extinction of small populations. *Am Nat* **146**: 489-518.

Macarthur JW. 1949. Selection for small and large mice in the house mouse. *Genetics* **34**: 194-209.

Mackay TFC, Fry JD, Lyman RF, Nuzhdin SV. 1994. Polygenic mutation in *Drosophila melanogaster*: estimates from response to selection of inbred strains. *Genetics* **136**: 937-951.

Maher GJ, McGowan SJ, Giannoulatou E, Verrill C, Goriely A, Wilkie AOM. 2016. Visualizing the origins of selfish de novo mutations in individual seminiferous tubules of human testes. *Proc Nat Acad Sci USA* **113**: 2454-2459.

Maule MM, Vizzini L, Merletti F, Magnani C, Pastore G, Richiardi L. 2007. Parental age and risk of acute lymphocytic leukaemia and embryonal tumours in the Piedmont Region, Italy. *Int J Epidemiol* **36**: 691-692.

McIntosh GC, Olshan AF, Baird PA. 1995. Paternal age and the risk of birth defects in offspring. *Epidemiology* **6**: 282-8.

McKee AM, Spear SF, Pierson TW. 2015. The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biol Conserv* **183**: 70–76.

Mendel JG. 1866. Experiments in plant hybridization. *J Royal Horticultural Soc* **26**: 1-32.

Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**: 1819-1829.

Misztal I, Legarra A, Aguilar I. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J Dairy Sci* **92**: 4648-4655.

Morgan A, Ness R, Keightley P, Colegrave N. 2014. Spontaneous mutation accumulation in multiple strains of the green alga, *Chlamydomonas reinhardtii*. *Evolution* **68**: 2589-2602.

Murray RK, Rodwell VW, Bender D, Botham KM, Weil PA, Kennelly PJ. 2012. *Harper's Illustrated Biochemistry*. McGraw Hill Professional, New York, NY, EUA.

Nazer J, Cifuentes L, Millán F, Vacaristas P, Köbrich S, Aguila A. 2008. La edad paterna como factor de riesgo para malformaciones congénitas. *Revista Médica de Chile* **136**: 201-208.

Nishi M, Nanjo K. 2011. Insulin gene mutations and diabetes. *J Diabetes Investig* **2**: 92-100.

Novo FJ. 2007. *Genética Humana. Conceptos, Mecanismos y Aplicaciones de la Genética en el Campo de la Biomedicina*. Prentice Hall, Madrid, España.

Nuzul W, Nielsen VH, Jensen J. 2014. Estimation of genetic variance components including mutation and epistasis using Bayesian approach in a selection experiment on body weight in mice. En *Proc 10th World Congr Genet Appl Livest Prod*, Vancouver, Canadá.

Ober C, Hyslop T, Hauck WW. 1999. Inbreeding effects on fertility in humans: evidence for reproductive compensation. *Am J Hum Genet* **64**: 225-231.

Okazaki R, Okazaki T, Sakabe K, Sugimoto K, Sugino A. 1968. Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc Nat Acad Sci USA* **59**: 598-605.

Ostensen T, Christensen OF, Henryon M, Nielsen B, Su G, Madsen P. 2011. Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genet Sel Evol* **43**: 38.

Pearson H. 2006. Genetics: what is a gene? *Nature* **441**: 398-401.

Peripato AC, de Brito RA, Matioli SR, Pletscher LS, Vaughn TT, Cheverud JJ. 2005. Epistasis affecting litter size in mice. *J Evol Biol* **17**: 593-602.

Quaas RL. 1976. Computing the diagonal elements and inverse of a large numerator

relationship matrix. *Biometrics* **32**: 949-953.

Quintanilla R, Varona L, Pujol R, Piedrafita J. 1999. Maternal animal model with correlation between maternal environmental effects of related dams. *J Anim Sci* **77**: 2904-2917.

Rothschild MF, Soller M. 1997. Candidate gene analysis to detect genes controlling traits of importance in domestic livestock. *Probe* **8**: 13-20.

Scheu A, Powell A, Bollongino R, Vigne J-D, Tresset A, Çakırlar C, Benecke N, Burger J. 2015. The genetic prehistory of domesticated cattle from their origin the the spread across Europe. *BMC Genetics* **16**: 54.

Shen LL, Sun HJ, Qian YM, Chen D, Zhan HX, Yang JG, Wang FL. Screening and identification of tobacco mutants resistant to tobacco and cucumber mosaic viruses. *J Agric Sci* (in press).

Shu J-J. 2017. A new integrated symmetrical table for genetic codes. *BioSystems* **151**: 21-26.

Simons MJ, Crow JF. 1997. Mutations affecting fitness in *Drosophila* populations. *Ann Rev Genom Hum Genet* **11**: 49-78.

Smith HF. 1936. A discriminant function of plant selection. *Ann Eugenics* **7**: 240-250.

Smith C, Bampton PR. 1977. Inheritance of reaction to halotance anesthiology in pigs. *Genet Res* **29**: 287-292.

Sorensen DA, Andersen S, Gianola D, Korsgaard I. 1995. Bayesian inference in threshold models using Gibbs sampling. *Genet Sel Evol* **27**: 229-249.

Straniero L, Rimoldi V, Soldà G, Mauri L, Manfredini E, Andreucci E, Bargiacchi S,

Penco S, Gesu GP, Del Longo A, Piozzi E, Asselta R, Primignani P. 2015. Two novel splicing mutations in the SLC45A2 gene cause oculocutaneous albinism type IV by unmasking cryptic splice sites. *J Hum Genet* **60**: 467-471.

Sugimoto K, Okazaki T, Okazaki R. 1968. Mechanism of DNA chain growth. II. Accumulation of newly synthesized short chains in E. coli infected with ligase-defective T4 phages. *Proc Nat Acad Sci USA* **60**: 1356-1362.

Swallow DM. 2003. Genetics of lactose persistence and lactose intolerance. *Annu Rev Genet* **37**: 197-219.

Taft R, Pang K, Mercer TR, Mattick JS. 2010. Non-coding RNAs: regulators of disease. *J Pathol* **220**: 126-139.

Vagnini L, Baruffi RL, Mauri AL, Petersen CG, Massaro FC, Pontes A, Oliveira JB, Franco JG Jr. 2007. The effects of male age on sperm DNA damage in an infertile population. *Reprod Biomed Online* **15**: 514-9

VanRaden PM. 2008. Efficient methods to compute genomic predictions. *J Dairy Sci* **91**: 4414-4423.

VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J Dairy Sci* **92**: 16-24.

Wade MJ. 1992. Sewall Wright: gene interaction and the shifting balance theory. *Oxford Surv Evol Biol* **8**: 33-62.

Watson JD, Crick FHC. 1953a. A structure for deoxyribose nucleic acid. *Nature* **171**: 737-738.

Watson JD, Crick FHC. 1953b. Genetical implications of the structure of

deoxyribonucleic acid. *Nature* **171**: 964-967.

Watson IR, Takahashi K, Futreal PA, Chin L. 2013. Emerging patterns of somatic mutations in cancer. *Nat Rev Genet* **14**: 703-718.

Wiecher L, Cuéllar H. 2007. Cambios fisiológicos durante el envejecimiento del sistema reproductor masculino. *Revista de Endocrinología y Nutrición* **15**: 207-216.

Wiggans GR, VanRaden PM, Cooper TA. 2011. The genomic evaluation system in the United States: past, present, future. *J Dairy Sci* **94**: 3202-3211.

Wilkins MHF, Stokes AR, Wilson HR. 1953. Molecular structure of deoxypentose nucleic acids. *Nature* **171**: 738-740.

Wray NR. 1990. Accounting for mutation effects in the additive genetic variance-covariance matrix and its inverse. *Biometrics* **46**: 177-186.

Wright S. 1922. Coefficients of inbreeding and relationship. *Am Nat* **56**: 330-338.

Yang Q, Wen SW, Leader A, Chen XK, Lipson J, Walker M. 2007. Paternal age and birth defects: how strong is the association? *Hum Reprod* **22**: 696-701.

Yoo BH. 1980. Long-term selection for a quantitative character in large replicate populations of *Drosophila melanogaster*. II. Lethal and visible mutants with large effects. *Genet Res* **35**: 19-31.

CAPÍTULO II

OBJETIVOS

Una vez planteados los antecedentes a la presente investigación se hace necesario establecer los objetivos que se pretenden cumplir con el desarrollo de este trabajo y que se señalan a continuación:

1. Desarrollar un modelo estadístico a partir de los modelos clásicos de evaluación genética que permita considerar el impacto de las nuevas mutaciones con efecto aditivo sobre el fenotipo, a la vez que la evolución de la variabilidad mutacional en función de la edad de los progenitores.
2. Evaluar el efecto de la edad y el sexo de los progenitores sobre la varianza aditiva mutacional del peso al destete en una línea isogénica de ratones C57BL/6J.
3. Evaluar el efecto de la edad y el sexo de los progenitores sobre la varianza aditiva mutacional del peso al nacimiento en los rebaños del núcleo de selección de la raza Bruna dels Pirineus.
4. Realizar un análisis comparativo de los resultados obtenidos sobre la variabilidad mutacional y el efecto de la edad y el sexo de los progenitores para ambas especies de mamíferos, así como sus consecuencias extrapolables a otras poblaciones de mamíferos.

CAPÍTULO III
MATERIALES I MÉTODOS

1. MODELO ANALÍTICO DESARROLLADO

Los dos trabajos de investigación que forman parte de la presente tesis doctoral fueron desarrollados con el objetivo de ejemplificar el comportamiento de la variabilidad mutacional aditiva en función de la edad y el sexo de los progenitores, trabajando con dos especies de mamíferos diferentes; por una parte se utilizó una línea isogénica de ratones de laboratorios y, por la otra, los registros productivos de una raza de vacuno cárnico, de importancia económica para el sector ganadero a nivel mundial. Cada fuente de datos se describirá detalladamente en posteriores secciones de esta tesis, así como las peculiaridades del enfoque analítico para cada una de ellas. No obstante, la parametrización general de modelo de análisis será la misma para ambas bases de datos y se describirá en detalle en esta sección.

1.1. Parametrización del modelo de análisis

En términos generales, el modelo mixto lineal completo se asumió con la siguiente estructura jerárquica:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_1\mathbf{p} + \mathbf{Z}_2\mathbf{a} + \mathbf{Z}_2\mathbf{m}_p + \mathbf{Z}_2\mathbf{m}_m + \mathbf{e}$$

donde \mathbf{y} era el vector de fenotipos, influidos por los correspondientes efectos sistemáticos (\mathbf{b}), ambientales permanentes (\mathbf{p}), genéticos aditivos (\mathbf{a}), y mutacionales de origen paterno (\mathbf{m}_p) y materno (\mathbf{m}_m). Finalmente, \mathbf{e} era el vector de residuos, y \mathbf{X} , \mathbf{Z}_1 y \mathbf{Z}_2 eran las matrices de incidencias respectivas. Centrándonos en los aspectos mutacionales de carácter aditivo, estos se diferenciaron de acuerdo al origen paterno (\mathbf{m}_p) o materno (\mathbf{m}_m) de los mismos, con el objetivo de comparar los efectos de ambas fuentes de nueva varianza mutacional aditiva, su magnitud y su evolución con la edad de los reproductores. Los efectos sistemáticos y ambientales permanentes fueron específicos para cada base de datos y se detallarán más adelante.

Para modelar correctamente los efectos mutacionales, se tomaron como punto de partida las ideas desarrolladas por Wray (1990) dentro del contexto de la herencia aditiva de nuevas mutaciones, modificándolas apropiadamente para el caso que nos ocupa.. En este contexto, la matriz global que incluye todas las relaciones mutacionales aditivas (\mathbf{M}) se puede generalizar como

$$\mathbf{M} = \mathbf{M}_1 + \mathbf{M}_2 + \dots + \mathbf{M}_n,$$

siendo n el número total de individuos de la genealogía, y \mathbf{M}_i la matriz de parentescos aditivos referentes a las nuevas mutaciones aparecidas en el i -ésimo individuo de la genealogía. Esto implica que \mathbf{M}_i tendrá únicamente ceros en las filas y columnas previas a i , y en las demás se calcularán los parentescos correspondientes condicionados a la parte de genoma heredada del individuo i ; nótese que Wray (1990) aplicaba los mismos procedimientos, aunque para simplificar, asumía generaciones no solapadas y calculaba \mathbf{M}_i para cada generación (en nuestro caso, al existir generaciones solapadas, resulta imprescindible expandir el procedimiento a todos los individuos analizados). La expansión de la matriz \mathbf{M} en n matrices de parentesco (\mathbf{M}_i) implicaría también la subdivisión de los efectos mutacionales aditivos (\mathbf{m}) en n vectores independientes de los efectos mutacionales aditivos parciales,

$$\mathbf{m} = \mathbf{m}_1 + \mathbf{m}_2 + \dots + \mathbf{m}_n,$$

capturando cada uno de ellos los efectos infinitesimales aditivos heredados a partir de las nuevas mutaciones ocurridas en el genoma del i -ésimo individuo (\mathbf{m}_i) y transmitidas a sus descendientes. También se puede observar que los efectos mutacionales fueron divididos en origen materno y origen paterno y fueron modelados a través de la distribución *a priori* de \mathbf{m}_p y \mathbf{m}_m . Tomando los efectos mutacionales específicos del padre surgidos en el genoma del individuo i th ($\mathbf{m}_{p,i}$), su distribución *a priori* fue establecida en una forma normal multivariante que será explicada en cada caso estudiado posteriormente.

1.2. Desarrollo Bayesiano

El modelo descrito anteriormente se desarrolló desde un punto de vista Bayesiano, asumiendo que la distribución posterior conjunta de todos los parámetros desconocidos era proporcional a la expresión que se detalla a continuación,

$$\begin{aligned}
 p(\mathbf{b}, \mathbf{p}, \sigma_p^2, \mathbf{a}, \sigma_a^2, \mathbf{m}_p, \sigma_{mp}^2, \lambda_p, \mathbf{m}_m, \sigma_{mm}^2, \lambda_m, \sigma_e^2 | \mathbf{y}) &\propto p(\mathbf{y} | \mathbf{b}, \mathbf{p}, \mathbf{a}, \mathbf{m}_p, \mathbf{m}_m, \sigma_e^2) \\
 &\times p(\mathbf{b}) p(\mathbf{p} | \sigma_p^2) p(\sigma_p^2) \\
 &\times p(\mathbf{a} | \mathbf{A}, \sigma_a^2) p(\sigma_a^2) \\
 &\times p(\mathbf{m}_p | \sigma_{mp}^2, \lambda_p, \mathbf{M}) p(\sigma_{mp}^2) p(\lambda_p) \\
 &\times p(\mathbf{m}_m | \sigma_{mm}^2, \lambda_m, \mathbf{M}) p(\sigma_{mm}^2) p(\lambda_m) \\
 &\times p(\sigma_e^2)
 \end{aligned}$$

donde \mathbf{A} era la matriz de parentescos aditivos (Wright, 1922), λ_p y λ_m eran los coeficientes de regresión para vincular la varianza mutacional paterna (σ_{mp}^2) y materna (σ_{mm}^2) con la edad del padre y de la madre en el momento de la concepción, y finalmente σ_p^2 , σ_a^2 y σ_e^2 correspondían a la varianzas ambiental permanente, genética aditiva, y residual, respectivamente. Se asumió una distribución normal multivariante (*NMV*) para los datos fenotípicos (\mathbf{y}),

$$p(\mathbf{y} | \mathbf{b}, \mathbf{p}, \mathbf{a}, \mathbf{m}_p, \mathbf{m}_m, \sigma_e^2) = NMV(\mathbf{Xb} + \mathbf{Z}_1\mathbf{p} + \mathbf{Z}_2\mathbf{a} + \mathbf{Z}_2\mathbf{m}_p + \mathbf{Z}_2\mathbf{m}_m, \mathbf{I}\sigma_e^2),$$

donde \mathbf{I} era una matriz de identidad con dimensión igual al número de elementos en \mathbf{y} . Paralelamente, para los efectos ambientales permanentes y genéticos aditivos también se asumieron distribuciones normales multivariantes,

$$p(\mathbf{p} | \sigma_p^2) = NMV(\mathbf{0}, \mathbf{I}\sigma_p^2) \text{ y } p(\mathbf{a} | \mathbf{A}, \sigma_a^2) = NMV(\mathbf{0}, \mathbf{A}\sigma_a^2).$$

En el caso de los efectos mutacionales, tanto paternos como maternos, fueron modelados a partir de la siguiente estructura normal multivariante (tomamos los efectos mutacionales paternos a modo de ejemplo),

$$p(\mathbf{m}_{p,i} | \mathbf{M}_i, \sigma_{mp}^2, \lambda_p) = NMV(\mathbf{0}, 0.5\mathbf{M}_i \sigma_{mp}^2 [1 + \varepsilon_{p,i} \lambda_p])$$

donde, $\varepsilon_{p,i}$ correspondía a la edad del padre en el momento de la concepción del i -ésimo individuo. Para los efectos mutacionales maternos se asumió la misma distribución, aunque incluyendo apropiadamente los parámetros $\mathbf{m}_{m,i}$, σ_{mm}^2 , λ_m y $\varepsilon_{m,i}$. Se asumieron *a priori*s planos para los demás parámetros del modelo (componentes de varianza, efectos fijos, y coeficientes de regresión sobre la edad de los progenitores).

1.3. Resolución del modelo Bayesiano

Dada la estructura Bayesiana descrita anteriormente, cualquier inferencia sobre los parámetros incluidos se realizó a partir de la distribución posterior conjunta. Dada la forma multidimensional de esta distribución posterior, la integración directa de la misma no resulta viable (Blasco, 2001). Esta problemática se resuelve a partir de la aplicación de las técnicas de cadenas de Márkov de Monte Carlo, que permiten obtener muestras aleatorias de los distintos parámetros del modelo a partir de sus distribuciones marginales posteriores. Para el caso que nos ocupa, se aplicaron procedimientos de muestreo de Gibbs (Geman y Geman, 1984; Gelfand y Smith, 1990; Gilks et al., 1996) para todos los parámetros (Sorensen et al., 1994; Casellas y Medrano, 2008), con la excepción de λ_p y λ_m que requirieron de procedimientos de muestreo de Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970). Para estos últimos parámetros, se partió de una distribución uniforme centrada en el último valor de muestreo de λ_p y λ_m , respectivamente, con límites en $\lambda_p \pm 0,1$ (o $\lambda_m \pm 0,1$). Esta función originó una tasa de aceptación de los muestreos superior al 15% en todos los análisis, alcanzando el 20% en alguno de ellos.

Cada modelo (ver sección siguiente) se analizó mediante tres cadenas de

Márkov de Monte Carlo independiente, con 500,000 iteraciones cada una, y después de descartar las primeras 10,000 *burn-in* (Raftery y Lewis, 1992). Dada la elevada autocorrelación existente entre muestras sucesivas, únicamente se tomó una de cada diez iteraciones para caracterizar la distribución marginal posterior de los distintos parámetros del modelo de análisis (50,000 iteraciones por cadena), asumiendo la propiedad ergódica de las cadenas (Gilks et al., 1996). Los análisis se realizaron mediante programas *ad hoc* escritos en lenguaje informático Fortran90, y se ejecutaron en un servidor con dos procesadores Intel® Xeon® E5-2400 con Linux CentOS 6.6.

1.4. Comparación de modelos

El modelo descrito anteriormente asume la inclusión de ambas fuentes de variabilidad mutacional (i.e., paterna y materna), aunque la relevancia estadística de estos efectos debería testarse apropiadamente. Además, cada una de estos factores del modelo se ve determinado por dos parámetros parcialmente independientes, el componente de varianza (σ_{mp}^2 y σ_{mm}^2) que determina si existe o no variabilidad mutacional aditiva, y el coeficiente de regresión (λ_p y λ_m) que evalúa si la varianza evoluciona con la edad de los progenitores. Los distintos modelos competitivos posibles fueron definidos mediante la eliminación de algunos (o todos) los parámetros relacionados con la variabilidad mutacional, tal como se detalla en la Tabla 1. Tanto las varianzas mutacionales como los coeficientes de regresión relacionados con la edad de los progenitores pueden ser excluidos del análisis fijándolos a cero. A modo de ejemplo, si $\sigma_{mp}^2 = 0$ se asume que el aporte mutacional paterno es nulo, mientras que si $\lambda_p = 0$ se descarta la posibilidad de que la varianza mutacional paterna se modifique con la edad del padre, independientemente de la estimación de σ_m^2 (siempre que $\sigma_m^2 > 0$). En total y con respecto a las contribuciones mutacionales paternas y maternas, se plantearon nueve modelos diferentes (Tabla 1).

Tabla 1. Definición de los modelos de análisis con base en los parámetros mutacionales estimados y fijados a cero.

	Parámetros mutacionales	
	Fijados a cero	Estimados
Modelo 0	$\sigma_{mp}^2, \sigma_{mm}^2, \lambda_p$ y λ_m	Ninguno
Modelo A	σ_{mm}^2, λ_p y λ_m	σ_{mp}^2
Modelo B	σ_{mm}^2 y λ_m	σ_{mp}^2 y λ_p
Modelo C	σ_{mp}^2, λ_p y λ_m	σ_{mm}^2
Modelo D	σ_{mp}^2 y λ_p	σ_{mm}^2 y λ_m
Modelo E	λ_p y λ_m	σ_{mp}^2 y σ_{mm}^2 ,
Modelo F	λ_m	$\sigma_{mp}^2, \sigma_{mm}^2$ y λ_p
Modelo G	λ_p	$\sigma_{mp}^2, \sigma_{mm}^2$ y λ_m
Modelo H	Ninguno	$\sigma_{mp}^2, \sigma_{mm}^2, \lambda_p$ y λ_m

Con la finalidad de dilucidar la relevancia biológica y estadística de cada modelo, se implementaron dos métodos estadísticos diferentes de comparación de modelos, el *deviance information criterion* (DIC; Spiegelhalter et al., 2002) y el factor de Bayes (FB; Kass y Raftery, 1995). El primero basa sus cálculos en el logaritmo promedio de la verosimilitud Bayesiana del mismo a lo largo de las distintas iteraciones, así como el logaritmo de la verosimilitud Bayesiana calculada a partir de la media posterior de los distintos parámetros del modelo (Spiegelhalter et al., 2002). Éste estadístico evalúa tanto el nivel de ajuste del modelo a los datos reales, como la complejidad del mismo en términos del número de parámetros a ser explicados. Al comparar modelos, resultarán preferibles aquellos con DIC más pequeños, indicando así un mejor ajuste y un menor grado de complejidad. En términos generales se consideran suficientes las diferencias de entre 3 y 5 unidades de DIC para considerar dos modelos como estadísticamente diferentes (Spiegelhalter et al., 2002).

Por otro lado, se calculó también el FB para comparar todos los modelos

por pares, partiendo del método desarrollado por Newton y Raftery (1994). Este estadístico Bayesiano calcula el cociente entre las probabilidades posteriores de los dos modelos a comparar (Kass y Raftery, 1995); así, un $FB > 1$ sugiere que el modelo de numerador resulta preferible ante el modelo del denominador, mientras que un $FB < 1$ favorece al modelo del denominador. La relevancia estadística inherente a los distintos FB se caracterizó mediante la aplicación de la escala de evidencias de Jeffreys (1984).

2. POBLACIONES Y DATOS ANALIZADOS

2.1. Ratones isogénicos

En el primer análisis se trabajó con los registros de peso al destete (tres semanas de edad) de una población de ratones C57BL/6J. La genealogía estaba compuesta por un total de 13.024 ratones, de los cuales 12.644 individuos disponían de fenotipo registrado para el análisis. Los animales se mantuvieron en las instalaciones del animalario de la Universidad de California (Davis, California, EUA) bajo estrictas condiciones ambientales y de manejo durante 46 generaciones no solapadas, desde octubre del año 1988 hasta mayo del año 2005. Más concretamente, esta línea de ratones fue fundada a partir de dos machos y seis hembras procedentes de *The Jackson Laboratory* (Bar Harbor, Maine, EUA). Cada año se producían entre dos y cinco generaciones, y cada generación implicaba de dos a 49 machos reproductores y de seis a 99 hembras reproductoras, con un promedio de 21,6 camadas por generación (Casellas y Medrano, 2008). El número de camadas por generación siempre fue muy variable, estando en función de la demanda de ratones para fines experimentales en la universidad. Para obtener cada nueva generación, se seleccionaban hembras y machos al azar de algunas pocas camadas de la generación anterior, favoreciéndose los cruzamientos entre hermanos completos.

Todas las acciones realizadas sobre los animales, protocolos e instalaciones seguían las directrices establecidas por la Asociación Internacional para la Evaluación y Acreditación del Cuidado de Animales de Laboratorio (AAALAC; <http://www.aaalac.org>). Estos ratones C57BL/6J fueron sometidos a condiciones ambientales controladas de temperatura ($21 \pm 2^\circ\text{C}$), humedad relativa (40 – 70 %) e iluminación (14 horas de luz y 10 horas de oscuridad), albergándolos en cajas de policarbonato específicamente diseñadas para el mantenimiento de esta especie de animales de laboratorio. Para la alimentación de los ratones se utilizó una dieta a base de alimento balanceado Purina 5008 (Ralston Purina, St. Louis, EUA; 23.5 % proteína, 6.5% grasa, 3.3 Kcal/g) y se les suministró agua *ad libitum*. Se registró de

manera sistemática toda la información relevante tanto desde un punto de vista reproductivo como de crecimiento de los ratones. Para cada camada se disponía de información referente al padre, la madre, la fecha de apareamiento, la fecha de nacimiento y el número de crías por parto (nacidos vivos y muertos). Además, las crías fueron pesadas de forma individual y marcadas en la oreja en el momento del destete (3 semanas después del nacimiento).

2.2. Vacuno cárnico

En el segundo análisis que se incluye en la presente tesis doctoral fueron utilizadas como fuentes de datos las explotaciones en control de rendimientos de la raza Bruna de los Pirineos. Se trata de una raza bovina de aptitud cárnica que se encuentra localizada mayoritariamente en las diferentes comarcas del Pirineo y el Prepirineo catalán (nordeste de España). Se trata de una población que tiene como principal característica la rusticidad, la longevidad y la alta habilidad materna. Esta raza es explotada habitualmente bajo sistemas extensivos de producción con trashumancia valle/montaña en épocas favorables; las explotaciones suelen ser de carácter intensivos en cuanto al engorde y finalizado de los terneros para el mercado (Sánchez-Belda, 2002). La raza Bruna de los Pirineos proviene de una antigua fusión entre el ecotipo local de bovinos de la zona Pirenaica de Catalunya con animales importados de la raza *Brown Swiss* procedente de Suiza (Fina, 2013). En sus inicios, la raza Bruna de los Pirineos presentaba una aptitud triple (i.e., carne, leche y trabajo), pero se ha ido especializando en la producción cárnica (Mujal, 1998). La Bruna de los Pirineos es una raza de tamaño mediano (toros de 950 a 1400 kg de peso vivo y vacas de 550 a 950 kg de peso vivo) con un censo aproximado de unas 30.000 vacas nodrizas; de estas, cerca de 12.000 están inscritas en el Libro Genealógico de la raza y aproximadamente 5.500 participan activamente en el Programa de Control de Rendimientos (Federació Catalana de la Raça Bruna dels

Pirineus (FEBRUPI), comunicación oral). Esta raza se explota en condiciones extensivas para la producción de terneros que suelen alcanzar un peso promedio en canal de 330 kg con 12.5 meses de edad (Serra et al., 2004).

El Programa de Mejora Genética de la raza se inició el año 1990 y se ha mantenido hasta la actualidad. Los objetivos principales de selección son, para las madres, mejorar la fertilidad, fomentar sus buenas cualidades maternas, y mantener las características de rusticidad inherentes a esta raza; para los terneros, se pretenden incrementar los pesos al destete, mejorando así los índices de crecimiento, dado que el producto que comercializan los ganaderos acostumbran a ser los terneros destetados (Fina, 2013). Además se procura obtener un peso medio al nacimiento que disminuya los índices de partos distócicos (Jordana y Pietrafita, 1993). Todo esto resulta posible gracias al Programa de Control de Rendimientos de la raza, un instrumento elemental para la gestión de la producción ganadera que permite aportar información muy valiosa al Libro Genealógico de la Bruna de los Pirineos, así como los registros productivos que representan el punto de partida para la evaluación genética y el mejoramiento sistemático de los animales. Los datos obtenidos permiten realizar un control productivo de las distintas explotaciones (Fina, 2013). Esto se debe a que los ganaderos inscritos en el programa de control de rendimientos deben llevar un control preciso de todos los sucesos productivos y reproductivos de su explotación ganadera. Un control eficaz implica el registro diario de todos los eventos o incidencias que suceden en cada ganadería tales como partos, pesajes, apareamientos e inseminaciones, entre otros. Toda esta información, debidamente informatizada, está centralizada por FEPRUBI, quién a su vez proporciona los resultados de las evaluaciones genéticas a los ganaderos, así como otras informaciones técnicas de interés y asesoramientos varios.

Para este estudio se tomaron los registros de peso al nacimiento de los terneros de raza pura nacidos entre los años 1985 y 2015 en 12 rebaños comerciales que participaban activamente en el Programa de Control de Rendimientos de la raza

Bruna de los Pirineos. Los rebaños fueron escogidos en base a la integridad de su genealogía ya que únicamente se incluyeron los fenotipos de terneros con padre y madre conocidos (Fina et al., 2012, 2013). Adicionalmente, cada rebaño debía disponer de conexión genealógica con los demás y aportar un mínimo de 100 fenotipos (Tarrés et al., 2010). Siguiendo las directrices de la *Beef Improvement Federation* (BIF, 1986), los terneros con pesos al nacimiento inferiores a 20 kg o superiores a 70 kg fueron descartados.

La base de datos final disponía de 8.130 registros de peso al nacimiento. La genealogía estaba compuesta por 10,266 animales, incluyendo 230 padres y 2.671 madres.

Para este estudio no se tramitó el permiso de la Comisión de Ética en la Experimentación Animal y Humana de la Universidad Autónoma de Barcelona (Bellaterra, España), porque los análisis se realizaron sobre datos pre-existentes (tanto los datos productivos como los reproductivos), y todos fueron tomados directamente por los mismos ganaderos de FEPRUBI dentro del contexto de su Programa de Control de Rendimientos. Esta toma de datos en ningún caso implicó la realización de actividades con los animales distintas al manejo estándar de una ganadería de vacuno cárnico en extensivo.

3. MODELOS OPERACIONALES

Aunque ambas bases de datos se analizaron bajo la misma estructura jerárquica del modelo mixto animal descrito anteriormente, la estructura de los efectos sistemáticos difirió claramente de una especie a la otra. En el caso del peso al destete de los ratones C57BL/6J, el elemento \mathbf{Xb} incluyó los efectos del sexo del ratón (macho o hembra), el tamaño de la camada (<5, 5 a 7, 8, 9, o >9 ratones) y la generación de nacimiento como efectos categóricos, y la edad al destete como efecto continuo. Por otro lado, el componente ambiental permanente ($\mathbf{Z_{ip}}$) capturaba el efecto camada (Casellas et al., 2014).

Para el análisis del peso al nacimiento en la raza Bruna de los Pirineos, los efectos sistemáticos considerados fueron todos discretos, incluyendo el sexo del ternero (macho o hembra), la edad de la madre al parto (2, 3, 4, 5, 6, o >6 años), el tipo de parto (simple o gemelar), y el efecto rebaño-año-estación (135 niveles distintos). En este caso, el efecto ambiental permanente se ajustó al efecto materno inherente a la madre del ternero, pudiendo capturar tanto influencias genéticas maternas como ambientales maternas.

BIBLIOGRAFÍA

- Beef Improvement Federation. 1986. *Beef Improvement Federation Guidelines*. North Carolina University Press, North Carolina, EUA.
- Blasco A. 2001. The Bayesian controversy. *J Anim Sci* **79**: 2023-2046.
- Casellas J, Gianola D, Medrano JF. 2014. Bayesian analysis of additive epistasis arising from new mutations in mice. *Genet Res* **96**: e008.
- Casellas J, Medrano JF. 2008. Within-generation mutation variance for litter size in inbred mice. *Genetics* **179**: 2147-2155.
- Fina M. 2013. Optimització de l'avaluació genètica de la raça bovina Bruna dels Pirineus. Tesis Doctoral, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), España.
- Fina M, Ibáñez-Escriche N, Piedrafita J, Casellas J. 2013. Canalization analysis of birth weight in Bruna dels Pirineus beef cattle. *J Anim Sci* **91**: 3070-3078.
- Fina M, Varona L, Piedrafita J, Casellas J. 2012. Sources of sire-specific genetic variance for birth and weaning weight in Bruna dels Pirineus beef calves. *Animal* **6**: 1931-1938.
- Gelfand A, Smith AFM. 1990. Sampling based approaches to calculating marginal densities. *J Am Stat Assoc* **85**: 398-409.
- Geman S, Geman D. 1984. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans Patt Anal Machine Intell* **6**: 721-741.
- Gilks WR, Richardson S, Spiegelhalter DJ. 1996. Introducing Markov chain Monte Carlo. En *Markov Chain Monte Carlo in Practice* (ed. Gilks WR, Richardson S, Spiegelhalter D). Chapman & Hall, Londres, Reino Unido, pp. 1-19.

- Hastings WK. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**: 97-109.
- Jeffreys H. 1984. *Theory of Probability*. Clarendon Press, Oxford, Reino Unido.
- Jordana J, Piedrafita J. 1993. Programa de mejora genética de la agrupación racial Bruna dels Pirineus. En *V Reunión Nacional de Mejora Genética Animal*. Consejería de Agricultura y Pesca, Junta de Andalucía, Córdoba, España.
- Kass RE, Raftery AE. 1995. Bayes factors. *J Am Stat Assoc* **90**: 773-795.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953. Equations of state calculations by fastcomputing machines. *J Chem Phys* **21**: 1087-1092.
- Mujal MM. 1998. Anàlisi demogràfica i genètica de la vaca Bruna dels Pirineus (Demographic and genetic analysis of the Bruna dels Pirineus beef breed. Tesis de Master, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), España.
- Newton MA, Raftery AE. 1994. Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *J Royal Stat Soc B* **56**: 1-48.
- Raftery AE, Lewis SM. 1992. How many iterations in the Gibbs sampler? En *Bayesian Statistics IV* (ed. Bernardo JM, Berger JO, Dawid AP, Smith AFM). Oxford University Press, Oxford, Reino Unido, pp. 763-773.
- Sanchez-Belda A. 2002. *Razas ganaderas españolas. I Bovinas*. FEAGAS y MAPAS, Madrid, España.
- Serra X, Gil M, Gispert M, Guerrero L, Oliver MA, Sañudo C, Campo MM, Panea B, Olleta JL, Quintanilla R, Piedrafita J. 2004. Characterization of young bulls of the Bruna dels Pirineus cattle breed (selected from old Brown Swiss) in relation to carcass, meat quality and biochemical traits. *Meat Sci* **66**: 425-436

- Sorensen DA, Wang CS, Jensen J, Gianola D. 1994. Bayesian analysis of genetic change due to selection using Gibbs sampling. *Genet Sel Evol* **26**: 333-360.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. 2002. Bayesian measures of model complexity and fit. *J Royal Stat Soc B* **64**: 583-639.
- Tarrés J, Fina M, Piedrafita J. 2010. Connectedness among herds of beef cattle bred under natural service. *Genet Sel Evol* **42**: 6-14.
- Wray NR. 1990. Accounting for mutation effects in the additive genetic variance-covariance matrix and its inverse. *Biometrics* **46**: 177-186.
- Wright S. 1922. Coefficients of inbreeding and relationship. *Am Nat* **56**: 330-338.

CAPÍTULO IV
RESULTADOS

1. DESARROLLO METODOLÓGICO

Aunque la presente tesis doctoral aborda el análisis de la varianza mutacional en dos especies distintas de mamíferos domésticos, resulta importante destacar el desarrollo metodológico que ha permitido precisamente la realización de estos estudios. Sin ánimo de menospreciar los análisis en sí mismos, la contribución que representa el modelo propuesto para abordar de manera novedosa la varianza mutacional aditiva debe destacarse como uno de los hitos más importantes de los detallados en el presente documento. Sobre la base de un modelo mixto lineal estándar para la evaluación genética de reproductores (Henderson, 1973), se han realizado dos contribuciones relevantes dentro del marco de trabajo de la varianza aditiva mutacional (Wray, 1990):

1. Inclusión de términos específicos para cada progenitor, diferenciando entre nuevas mutaciones con origen paterno ($\mathbf{Z}_2\mathbf{m}_p$) y materno ($\mathbf{Z}_2\mathbf{m}_m$). Esto implicó un replanteamiento importante en la construcción de la matriz de covarianzas aditivas mutacionales (\mathbf{M}), pasando del diseño centrado en generaciones que proponía originalmente Wray (1990), a una estructura más flexible y específica de cada individuo. Esta modificación, en ningún caso altera las adaptaciones propuestas por Casellas y Medrano (2008) para el análisis eficiente de los efectos mutacionales aditivos dentro de un contexto Bayesiano, sino que permite su utilización de manera sistemática y sin requerimientos adicionales.
2. Re-parametrización de la varianza aditiva mutacional paterna (σ_{mp}^2) y materna (σ_{mm}^2) con un segundo nivel de jerarquización que considera la posibilidad de cambios lineales con la edad del progenitor (i.e., $\sigma_{mp}^2[1+\varepsilon_{p,i}\lambda_p]$). Resulta realmente importante destacar que esta estructura no asume forzosamente la presencia de una varianza mutacional creciente con la edad del progenitor (compatible con los resultados sugeridos en la especie humana (Goriely y

Wilkie, 2012)), sino que permite cualquier escenario posible, incluyendo tanto la ausencia de varianza mutacional ($\sigma_{mp}^2 = 0$), como la presencia de varianza mutacional invariable con la edad de los progenitores ($\lambda_p = 0$), o la reducción en la varianza mutacional a medida que envejecen los progenitores ($\sigma_{mp}^2 > 0$ y $\lambda_p < 0$).

Los requerimientos para la utilización de esta aproximación se limitan a la disponibilidad de registros fenotípicos y genealógicos fiables. Esto permitiría su implementación en la mayoría de especies domésticas, e incluso en humanos, sin necesidad de generar registros adicionales mediante tecnologías costosas como los genotipados masivos o la secuenciación (Wetterstrand, 2017).

2. ANÁLISIS SOBRE ESPECIES DOMÉSTICAS

2.1. Ratones C57BL/6J

Se estudió la contribución de las nuevas mutaciones aditivas sobre el peso al destete de ratones isogénicos. El componente genético se compartimentó en dos fuentes de variación diferentes, la variabilidad genética aditiva presente en la generación de fundadores (Henderson, 1973), y la varianza aditiva *de novo* originada por mutaciones en las subsiguientes generaciones (Wray, 1990; Casellas y Medrano, 2008; Casellas et al., 2010). Además, el componente mutacional se jerarquizó en dos términos independientes en función de si tenían origen paterno o materno, además de permitir que su varianza evolucionara de manera lineal con la edad de los progenitores. Esto implicaba un total de nueve modelos posibles distintos (Tabla 1) que se compararon mediante dos estadísticos distintos, el DIC (Spiegelhalter et al., 2002) y el factor de Bayes (Kass y Raftery, 1995).

El DIC más bajo se obtuvo con el modelo F (14.992,4), con una diferencia de casi 10 unidades con el siguiente modelo con mejor ajuste (modelo E; 15.002,2). Tal como se detalla en la Tabla 2, estos valores resultaron de promediar las tres cadenas de Márkov de Monte Carlo que se lanzaron para el análisis de cada modelo. Resulta importante destacar que el DIC mostró una variabilidad mínima para cada modelo, con un error típico que siempre se mantuvo por debajo de 1. En cualquier caso, el modelo con mejor ajuste era el que incluía varianza mutacional tanto paterna como materna y, además, la varianza mutacional paterna evolucionaba con la edad del progenitor. De hecho, el modelo E (segundo en cuanto a mejor ajuste), era una simplificación del modelo F, descartando el coeficiente de regresión para el componente mutacional paterno ($\lambda_p = 0$). La importancia de los parámetros σ_{mp}^2 , σ_{mm}^2 y λ_p quedaba fuera de toda duda con estos resultados, resaltándose aún más su relevancia si consideramos que el modelo 0 (excluía todos los parámetros mutaciones) fue uno de los que alcanzó peores ajuste de todos, únicamente superado

por el modelo C.

Tabla 2. Comparación de la bondad de ajuste evaluada mediante el *deviance information criterion* (Spiegelhalter et al., 2012) para los nueve modelos posibles ante el análisis del peso al destete en ratones C57BL/6J.

	Parámetros mutacionales ¹		<i>Deviance information criterion</i> ²	
	Fijados a cero	Estimados	Media	Error típico
Modelo 0	$\sigma_{mp}^2, \sigma_{mm}^2, \lambda_p$ y λ_m	Ninguno	15.054,4	0,9
Modelo A	σ_{mm}^2, λ_p y λ_m	σ_{mp}^2	15.035,2	0,7
Modelo B	σ_{mm}^2 y λ_m	σ_{mp}^2 y λ_p	15.014,7	0,9
Modelo C	σ_{mp}^2, λ_p y λ_m	σ_{mm}^2	15.063,0	0,8
Modelo D	σ_{mp}^2 y λ_p	σ_{mm}^2 y λ_m	15.051,0	0,8
Modelo E	λ_p y λ_m	σ_{mp}^2 y σ_{mm}^2	15.002,2	0,6
Modelo F	λ_m	$\sigma_{mp}^2, \sigma_{mm}^2$ y λ_p	14.992,4	0,9
Modelo G	λ_p	$\sigma_{mp}^2, \sigma_{mm}^2$ y λ_m	15.010,7	0,6
Modelo H	Ninguno	$\sigma_{mp}^2, \sigma_{mm}^2, \lambda_p$ y λ_m	15.009,1	0,9

¹Varianza mutacional paterna (σ_{mp}^2) y materna (σ_{mm}^2), así como los coeficientes de regresión sobre la edad del padre (λ_p) y la madre (λ_m).

²Media de las estimas obtenidas a partir de tres cadenas de Márkov de Monte Carlo independientes.

Los resultados obtenidos con el factor de Bayes corroboraron la superioridad del modelo F. Al comparar este modelo (como numerador) con cada uno de los demás se alcanzaron factores de Bayes superiores a 1 en todos los casos y fluctuando entre 10,99 (modelo E) y 1.111,1 (modelo C). Estos valores implican evidencias a favor del modelo F entre fuertes ($FB > 10$) y decisivas ($FB > 100$) según la escala descrita por Jeffreys (1961). Los resultados obtenidos con los modelos G y H descartan la posibilidad de detectar cambios relevantes de la varianza mutacional materna con la edad de la madre en esta base de datos, mientras que la evolución de la varianza mutacional paterna se corrobora por el modelo F. En cualquier caso, la consistencia observada entre el DIC y el factor de Bayes permite corroborar los

resultados y las conclusiones obtenidas a partir de los análisis realizados.

Tabla 3. Factor de Bayes entre cada combinación de modelos de análisis; el cociente considera como numerador el modelo especificado en cada columna, y como denominador el modelo detallado en cada fila.

	0	A	B	C	D	E	F	G	H
0	-	2,26	4,72	0,23	1,33	23,33	256,41	8,46	10,51
A	0,44	-	2,09	0,10	0,59	10,34	113,64	3,75	4,66
B	0,21	0,48	-	0,05	0,28	4,95	54,35	1,79	2,23
C	4,33	9,78	20,44	-	5,78	101,11	1.111,1	36,67	45,56
D	0,75	1,69	3,54	0,17	-	17,50	192,31	6,35	7,88
E	0,04	0,10	0,20	0,01	0,06	-	10,99	0,36	0,45
F	0,00	0,01	0,02	0,00	0,01	0,09	-	0,03	0,04
G	0,12	0,27	0,56	0,03	0,16	2,76	30,30	-	1,24
H	0,10	0,21	0,45	0,02	0,13	2,22	24,39	0,80	-

A continuación, se describen los componentes de varianza obtenidos a partir del modelo con mejor ajuste (modelo F). Los resultados detallados en la Tabla 4 evidenciaron que la varianza genética aditiva era importante en la población de ratones C57BL/6J, y debe verse como un resultado sorprendente dado que se trata de una línea isogénica de ratones. Mientras la varianza genética aditiva (σ_a^2) representaba aproximadamente un 10% de la variabilidad total, las varianzas mutacionales aditivas rozaban un 4%, incrementándose con la edad del progenitor paterno. La media posterior para la varianza mutacional paterna (σ_{mp}^2) resultó netamente superior a la materna (Tabla 4), aunque los intervalos de credibilidad se solapaban, y esto descartaba la posibilidad de sugerir diferencia estadísticas relevantes entre ambas fuentes de variación. No obstante, donde sí se diferenciaban estos componentes de varianza era en la evolución de los mismos a medida que

envejecían los progenitores. La varianza mutacional materna era estable con el tiempo, mientras que la paterna se modificaba a partir de un coeficiente de regresión que implicaba un incremento de 0,002 unidades al cuadrado por día. Este coeficiente podría parecer pequeño, pero su intervalo de credibilidad descarta claramente el valor nulo, e implica un incremento en la varianza desde 0,128 g² (pubertad) a 0,185 g² (machos de 1 año de edad) durante el primer año de vida del progenitor (~45%).

Tabla 4. Componentes de varianza y parámetros relacionados obtenidos en el análisis del peso al destete en ratones C57BL/6J.

Parámetro ¹	Media Posterior	Percentiles (2,5% a 97,5%)
σ_a^2	0,413	0,191 a 0,590
σ_p^2	1,498	0,867 a 2,152
σ_e^2	1,987	1,725 a 2,266
σ_{mp}^2	0,117	0,064 a 0,170
λ_p	0,002	$5,3 \times 10^{-5}$ a 0,049
σ_{mm}^2	0,054	0,018 a 0,105

¹Varianzas genética aditiva (σ_a^2), ambiental permanente (σ_p^2), residual (σ_e^2), mutacional paterna (σ_{mp}^2), y mutacional materna (σ_{mm}^2), más el coeficiente de regresión para la parametrización de la varianza mutacional paterna en función de la edad del padre (λ_p).

2.2. Terneros de raza Bruna de los Pirineos

Los análisis se replicaron sobre una raza doméstica, la Bruna de los Pirineos, y sobre un carácter implicado en su programa de mejora genética (Fina, 2013), relacionado con el desarrollo corporal de los individuos, al igual que en el ejemplo anterior. No obstante, este estudio se centró en el peso al nacimiento, avanzándose

claramente el momento fisiológico en el que se registraron los fenotipos. Al analizar el ajuste estadístico de los nueve modelos competitivos posibles, nuevamente el modelo F alcanzó el mejor ajuste con el DIC más bajo (2.698.013), y con una diferencia de casi 2.000 unidades con el siguiente (modelo E). En este caso, el peor ajuste lo alcanzó el modelo 0, precisamente el que excluye todos los parámetros relacionados con el componente mutacional (DIC = 2.709.255). Dentro de este contexto, la relevancia estadística de las varianzas mutacional paterna y materna, así como el coeficiente de regresión para la edad del padre se pueden considerar corroboradas. Nuevamente se descartó el coeficiente de regresión para la edad de la madre (Tabla 5), sugiriéndose un comportamiento muy similar entre ambas especies de mamíferos.

Tabla 5. Comparación de la bondad de ajuste evaluada mediante el *deviance information criterion* (Spiegelhalter et al., 2012) para los nueve modelos posibles ante el análisis del peso al nacimiento en terneros de raza Bruna de los Pirineos.

	Parámetros mutacionales ¹		<i>Deviance information criterion</i> ²	
	Fijados a cero	Estimados	Media	Error típico
Modelo 0	σ_{mp}^2 , σ_{mm}^2 , λ_p y λ_m	Ninguno	2.709.255	789
Modelo A	σ_{mm}^2 , λ_p y λ_m	σ_{mp}^2	2.705.547	691
Modelo B	σ_{mm}^2 y λ_m	σ_{mp}^2 y λ_p	2.702.225	804
Modelo C	σ_{mp}^2 , λ_p y λ_m	σ_{mm}^2	2.708.761	659
Modelo D	σ_{mp}^2 y λ_p	σ_{mm}^2 y λ_m	2.705.775	701
Modelo E	λ_p y λ_m	σ_{mp}^2 y σ_{mm}^2	2.699.983	664
Modelo F	λ_m	σ_{mp}^2 , σ_{mm}^2 y λ_p	2.698.013	554
Modelo G	λ_p	σ_{mp}^2 , σ_{mm}^2 y λ_m	2.701.983	634
Modelo H	Ninguno	σ_{mp}^2 , σ_{mm}^2 , λ_p y λ_m	2.700.143	701

¹Varianza mutacional paterna (σ_{mp}^2) y materna (σ_{mm}^2), así como los coeficientes de regresión sobre la edad del padre (λ_p) y la madre (λ_m).

²Media de las estimas obtenidas a partir de tres cadenas de Márkov de Monte Carlo

independientes.

La comparación de los modelos mediante factores de Bayes no arrojó discrepancias en relación con las conclusiones obtenidas con el DIC. En este caso, el modelo F reveló una superioridad manifiesta ante los demás modelos, con un factor de Bayes (modelo F como numerador) de entre 83,41 y 6.251,0 (evidencias entre muy fuertes y decisivas según la escala de Jeffreys (1961)). También el modelo 0 resultó inferior a todos los demás, evidenciándose la necesidad de considerar los componentes de varianza mutacional al analizar este carácter productivo en la raza Bruna de los Pirineos.

Tabla 6. Factor de Bayes entre cada combinación de modelos de análisis; el cociente considera como numerador el modelo especificado en cada columna, y como denominador el modelo detallado en cada fila.

	0	A	B	C	D	E	F	G	H
0	-	19,38	41,88	1,88	10,63	18,75	6.251,0	68,75	75,01
A	0,05	-	2,16	0,10	0,55	0,97	322,58	3,55	3,87
B	0,02	0,46	-	0,04	0,25	0,45	149,25	1,64	1,79
C	0,53	10,33	22,33	-	5,67	10,02	3.333,1	36,67	40,05
D	0,09	1,82	3,94	0,18	-	1,76	588,24	6,47	7,06
E	0,05	1,03	2,23	0,10	0,57	-	313,2	3,67	4,07
F	0,00	0,00	0,01	0,00	0,00	0,00	-	0,01	0,01
G	0,01	0,28	0,61	0,03	0,15	0,27	90,91	-	1,09
H	0,01	0,26	0,56	0,03	0,14	0,25	83,41	0,92	-

Centrándonos en el modelo F, en la Tabla 7 se reportan las estimas y los intervalos de credibilidad de los componentes de varianza y el coeficiente de regresión para la varianza mutacional paterna. La varianza mayoritaria fue la residual

(14,202 kg²), seguida por la genética aditiva (6,623 kg²) y la ambiental permanente (1,672 kg²). A continuación, cerraban la lista los componentes de varianza mutacional tanto paterna (0,591 kg²) como materna (0,308 kg²), con valores medios ligeramente superiores para el primero, aunque los intervalos de credibilidad se solapaban claramente; dentro de este contexto, no se podría hablar de diferencias estadísticas entre ambos sexos. Paralelamente, el coeficiente de regresión evidenció un incremento pequeño pero relevante en la varianza mutacional paterna, de 0,001 kg² por día. Al igual que en el caso de los ratones C57BL/6J, debemos considerar este coeficiente de regresión como estadísticamente relevante, ya que su intervalo de credibilidad excluye el valor nulo ($2,7 \times 10^{-4}$ a 0,013). Nótese que la distribución *a priori* para este parámetro permitía ese valor, dado que se asumió una función uniforme entre $-\infty$ y ∞ , con mínima informatividad (Blasco, 2001)

Tabla 7. Componentes de varianza y parámetros relacionados obtenidos en el análisis del peso al nacimiento en terneros de raza Bruna de los Pirineos.

Parámetro ¹	Media Posterior	Percentiles (2,5% a 97,5%)
σ_a^2	6,623	5,014 a 8,511
σ_p^2	1,672	0,904 a 2,338
σ_e^2	14,202	13,107 a 15,373
σ_{mp}^2	0,591	0,165 a 0,944
λ_p	0,001	$2,7 \times 10^{-4}$ a 0,013
σ_{mm}^2	0,308	0,140 a 0,477

¹Varianzas genética aditiva (σ_a^2), ambiental permanente (σ_p^2), residual (σ_e^2), mutacional paterna (σ_{mp}^2), y mutacional materna (σ_{mm}^2), más el coeficiente de regresión para la parametrización de la varianza mutacional paterna en función de la edad del padre (λ_p).

BIBLIOGRAFÍA

- Blasco A. 2001. The Bayesian controversy in animal breeding. *J Anim Sci* **79**: 2023-2046.
- Casellas J, Caja G, Piedrafita J. 2010. Accounting for arritive genetic mutations on litter size in Ripollesa sheep. *J Anim Sci* **88**: 1248-1255.
- Casellas J, Medrano JF. 2008. Within-generation mutation variance for litter size in inbred mice. *Genetics* **179**: 2147-2155.
- Fina M. 2013. *Optimització de l'avaluació genètica de la raça bovina Bruna dels Pirineus*. Tesis Doctoral, Universitat Autònoma de Barcelona, Bellaterra, Barcelona. España.
- Goriely A, Wilkie AO. 2012. Paternal age effect mutations and selfish spermatogonial selection: causes and consequences for human disease. *Am J Hum Genet* **90**: 175-200.
- Henderson CR. 1973. Sire evaluations and genetic trends. In *Proc Anim Breed Genet Symp in Honor of Jay L Lush*. American Society of Animal Science and American Dairy Science Association, Champaign, Illinois, USA.
- Jeffreys H. 1961. *Theory of Probability*. Clarendon Press, Oxford, Reino Unido.
- Kass RE, Raftery AE. 1995. Bayes factors. *J Am Stat Assoc* **90**: 773-795.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. 2002. Bayesian measures of model complexity and fit. *J Royal Stat Soc B* **64**: 583-639.
- Wetterstrand KA. 2017. Sequencing costs: Data from the NHGRI Genome Sequencing Program (GSP). Disponible en www.genome.gov/sequencingcostsdata. Accedido el 21 de mayo de 2017.

Wray NR. 1990. Accounting for mutation effects in the additive genetic variance-covariance matrix and its inverse. *Biometrics* **46**: 177-186.

CAPÍTULO V

DISCUSIÓN

1. METODOLOGÍA Y CONTEXTO DE ANÁLISIS

Aunque ha suscitado un interés creciente durante las últimas décadas (Lynch, 1988; Lynch et al., 1995; Houle et al., 1996; Casellas y Medrano, 2008), todavía es mucho lo que falta por investigar sobre la contribución de las nuevas mutaciones en el proceso evolutivo de las especies y sus contribuciones potenciales a los planes de mejora genética del ganado y los cultivos. Es por ello que el desarrollo de metodologías que permitan un abordaje fino de este fenómeno biológico, debe verse como una contribución muy relevante para múltiples campos de la ciencia, especialmente dentro de la genética. Dentro de este contexto, la presente tesis doctoral se centra en un aspecto muy concreto de las mutaciones en mamíferos, no tanto su contribución puntual sobre la variabilidad genética de la población, sino la evolución de la misma a medida que envejecen los individuos reproductores. La nueva parametrización que se ha desarrollado en el marco metodológico de la inferencia Bayesiana (Blasco, 2001), tiene como finalidad describir el patrón lineal de incremento o decrecimiento de la variabilidad genética originada a partir de nuevas mutaciones aditivas en cada progenitor en particular, así como también la caracterizar su acumulación a lo largo de toda la vida reproductiva. Sin duda representa un punto de partida muy importante para futuras investigaciones centradas en este ámbito y su aplicación potencial tanto en especies ganaderas como de laboratorio.

Tal como se puede comprobar en la literatura científica disponible, la carga mutacional se ha venido asumiendo como un aporte constante y homogéneo a lo largo de la vida reproductiva (Lynch, 1988; Houle et al., 1996). De hecho, y centrándonos en características como el peso corporal y el tamaño de la camada en poblaciones de especies de mamíferos, la invariabilidad de la tasa de contribución de nuevas mutaciones se asume por sistema (Keightley y Hill 1992; Casellas et al. 2014; Nuzul et al. 2014), incluso cuando se pretende modelar junto con otros fenómenos biológicos más complejos, como la epistaxis. Es necesario resaltar que la

parametrización desarrollada en este trabajo de investigación resulta pionera en cuanto a metodología estadística, y ofrece un marco de trabajo más flexible para evaluar la evolución de la varianza mutacional aditiva en función de la edad de los progenitores.

Como punto de partida se asumió un fraccionamiento dicotómico de la varianza genética aditiva, asumiendo dos componentes básicos y fundamentales. En primer lugar la varianza genética aditiva presente en la generación de individuos fundadores, representando esta la variabilidad genética ancestral existente en la población junto con la contribución, si cabe, de las mutaciones aditivas acaecidas en las generaciones precedentes (Henderson, 1973). Paralelamente, se consideró también un componente de varianza mutacional, siendo este el resultado de las mutaciones aditivas ocurridas en los animales que conforman la genealogía disponible (Lynch, 1988; Wray, 1990). Esta partición no difería de la asumida en estudios precedentes (Casellas y Medrano, 2008; Casellas et al., 2010), pero daba pie al siguiente nivel de jerarquización, siendo este el que realmente permitía una caracterización más fina de la variabilidad mutacional aditiva. El componente de varianza mutacional se dividía nuevamente en dos parámetros independientes, capturando cada uno de ellos la contribución en términos de nuevas mutaciones de cada uno de los progenitores; además, el modelo de análisis desarrollado permitía que estas varianzas mutacionales paterna y materna evolucionaran con la edad de cada progenitor, de manera lineal, aunque abriendo la puerta también a otras funciones estadísticas más flexibles (e.g., cuadrática, exponencial, ...) en caso de ser necesario.

El procedimiento desarrollado se fundamenta en los planteamientos algebraicos y analíticos desarrollado por Wray (1990) y Casellas y Medrano (2008). De hecho, las fechas de publicación de ambos trabajos (i.e., casi tres y una década atrás, respectivamente) destacan el lento avance que se ha realizado en el ámbito de la varianza mutacional durante la últimas décadas. El trabajo de Wray (1990)

describía el procedimiento para construir la matriz de parentescos mutacionales, siendo esta una estructura básica para determinar la varianza mutacional en sí. Los planteamientos originales de Wray (1990) asumían un escenario claro donde las nuevas mutaciones, con efectos pequeños y aditivos en el fenotipo, surgían de manera independientemente en cualquier individuo tanto fundador como de las demás generaciones de la genealogía. Estas nuevas mutaciones podían ser heredadas naturalmente por sus descendientes siguiendo un modelo clásico aditivos e infinitesimales y, como consecuencia, permitían el cálculo de parentesco siguiendo los procedimientos descritos por Wright (1922), con ligeras adaptaciones. De hecho, Wray (1990) simplificaba el cálculo asumiendo generaciones no solapadas de individuos. Esto permitía calculando de manera independiente el parentesco inherente a las mutaciones acaecidas en una generación, para los individuos de esa generación y sus descendientes. No obstante, los planteamientos de Wray (1990) se pueden expandir fácilmente a poblaciones con generaciones solapadas, reduciendo el tamaño de cada “generación” a un único individuo, y calculando, por lo tanto, tantas matrices de parentesco como individuos. En conjunto, esta parametrización proveía de una herramienta muy prometedora para dilucidar los aspectos relevantes e inherentes a las nuevas mutaciones, sus patrones de emergencia, y sus contribuciones esperadas tanto sobre las siguientes generaciones como para la varianza genética total.

En segundo lugar, los procedimientos desarrollados en la presente tesis doctoral se fundamentan sobre los planteamientos descritos por Casellas y Medrano (2008). En este caso, y asumiendo la estructura de matrices de parentesco mutacional previsto por Wray (1990), los autores simplificaron los procedimientos para inferir la magnitud de la varianza mutacional dentro de un contexto puramente Bayesiano. Esto implicaba escindir la varianza genética total en varianza genética aditiva de la población de fundadores, y varianza mutacional derivada de las nuevas mutaciones, apareciendo esta tanto para fundadores como no fundadores. Aunque esta estructuración resulta quizá obvia actualmente, permitió en su momento simplificar

los procedimientos de computación. Al generar dos componentes de varianza independientes y claramente identificables en el modelo, toda la inferencia se realizaba a partir de muestreo de Gibbs (Gelfand y Smith, 1990) en lugar de implementar abordajes genéricos y claramente menos eficientes como el de Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970).

En conjunto, el método desarrollado generaliza Wray (1990) y Casellas y Medrano (2008), dando un paso más para una correcta caracterización del fenómeno mutacional. Debe destacarse que esa metodología es sumamente novedosa y permite abordar de manera completamente flexible cualquier base de datos, requiriéndose únicamente información fenotípica y genealógica. De hecho, es precisamente esta flexibilidad la que debe tenerse en cuenta en el momento de utilizar el modelo desarrollado. Aunque la forma más completa considera tanto varianza mutacional paterna como materna, así como que ambas evolucionan con la edad de los progenitores pero de manera independiente, no implica necesariamente que este sea el mejor modelo posible. Para dilucidarlo, no ha bastado con el desarrollo del modelo, sino que ha sido necesario implementar procedimientos específicos para la comparación de modelos. Más concretamente, los procedimientos de computación se han completado con el cálculo del DIC (Spiegelhalter et al., 2002) y el factor de Bayes (Kass y Raftery, 1995); ambos métodos evalúan la relevancia tanto biológica como estadística de cada modelo en términos de su grado de complejidad (i.e., número de parámetros del modelo) y/o ajuste sobre los datos. El estadístico DIC (Spiegelhalter et al., 2002) se fundamenta en el cálculo del logaritmo de la verosimilitud, tanto puntualmente en el procedimiento de muestreo como sobre la media posterior de los parámetros del modelo, y favorece aquellos modelos con valores más bajos. En términos generales se consideran como “estadísticamente relevantes” las diferencias a partir de entre 3 a 5 unidades DIC (Spiegelhalter et al. 2002). Paralelamente, el factor de Bayes permite comparar la probabilidad posterior de dos modelos mediante un simple cociente, y en este caso se aplicaron específicamente los algoritmos descritos por Newton y Raftery (1994) para el cálculo

de las probabilidades implicadas. Dado que el factor de Bayes se calcula como un cociente de probabilidades entre dos modelos competitivos (Kass y Raftery, 1995), el resultado se interpreta precisamente en relación al valor de referencia que representa el 1. Así, un factor de Bayes mayor que 1 sugiere que el modelo situado como numerador resulta preferible en relación con modelo del denominador, mientras que un factor de Bayes menor que 1 favorece el modelo del denominador. La magnitud de la ventaja/desventaja de cada modelo dependerá de la diferencia en relación con valor de referencia, 1. Esta relevancia estadística se puede categorizar y estandarizar mediante la aplicación de la escala de evidencias descrita por Jeffreys (1961) hace ya más de medio siglo.

Las ventajas de la metodología desarrollada deben verse no únicamente en la capacidad de modelar y explicar de manera sumamente flexible un fenómeno complejo como la mutación y su evolución a lo largo de la edad de los progenitores, sino también en la posibilidad de adaptar métodos estadísticos estándar para comparar modelos y determinar el que alcanzar un mejor ajuste. La utilización del DIC (Spiegelhalter et al., 2002) y el factor de Bayes (Kass and Raftery 1995) simplifica sobremano la evaluación de cada modelo, sin necesidad de establecer el comportamiento estadístico de cada método de comparación, dado que este se ha evaluado repetidamente en la literatura científica y es bien conocido.

A modo de ejemplo, el modelo desarrollado se ha aplicado sobre dos especies de mamíferos domésticos, una dentro del ámbito de los animales de laboratorio y la otra claramente encuadrada en los sistemas ganaderos de producción. Tanto los análisis realizados sobre ratones isogénicos C57BL/6J como sobre terneros de raza Bruna de los Pirineos, deben verse como ejemplos relevantes de la flexibilidad del modelo. Esta flexibilidad no se centra únicamente en las características y las peculiaridades del componente mutacional y su evolución a lo largo de la edad de los progenitores, sino también en la capacidad de adaptarse a distintas especies animales. Las diferencias fenotípicas/ambientales entre vacas y ratones quedan fuera de toda

duda, pero esto se ve fácilmente acomodado dentro de la estructura de modelos mixtos (Henderson, 1973). Así, no resulta atrevido aseverar que la metodología desarrollada puede aplicarse a casi cualquier especie doméstica, adaptando las peculiaridades de la misma dentro de la estructura de efectos sistemáticos y permanentes del modelo mixto, y caracterizando sus componentes genéticos tal como se ha descrito anteriormente en el apartado de “materiales y métodos” de la presente tesis doctoral.

2. VARIANZA MUTACIONAL EN RATONES ISOGÉNICOS C57BL/6J

Muchos investigadores consideran al ratón de laboratorio como un animal casi perfecto desde el punto de vista experimental. Justo a su talla pequeña, su intervalo generacional corto, su fácil mantenimiento, y su gran rendimiento reproductivo a largo de todo el año (Festing, 1998), sus características genéticas, cuando se las considera en conjunto, lo hacen un modelo único para la genética experimental. La posibilidad de contar con líneas de roedores que estén genéticamente definidas y homogéneas ha sido uno de los principales avances de la ciencia de los animales de laboratorio. Las líneas consanguíneas son el prototipo de las líneas genéticamente estandarizadas o isogénicas, debido a que su constitución genética está fijada en forma casi definitiva. Hace más de 100 años que se cuenta con líneas consanguíneas de ratones, ratas y cobayos, todas de gran valor porque permiten hacer experimentos eliminando la variabilidad de origen genético (Festing, 1996). O al menos esto es lo que se pretendía...

En términos generales, se ha venido considerando que las principales características de las líneas consanguíneas debían ser (1) la isogenicidad, (2) la homocigosis, (3) la individualidad, (4) la uniformidad fenotípica, y (5) la estabilidad genética a largo plazo (Festing, 1996, 1998). En ningún momento se deseaba o consideraba la participación de la mutación, dado que este es un cambio natural, universal, aleatorio e inevitable, que haría inviable alcanzar de forma completa ninguna de las cinco características enumeradas anteriormente. De hecho, la mayoría de las líneas isogénicas de ratones de laboratorio se han generado mediante cruzamientos consanguíneos, intentando alcanzar la homogeneidad genética mediante el incremento sistemático de la consanguinidad hasta su máximo teórico del 100%. La mutación debe verse como un mecanismo genético opuesto a la consanguinidad, generador de diversidad genética y variabilidad mediante la introducción de nuevos alelos. En definitiva, como el elemento antagónico a los objetivos inherentes a la creación de líneas isogénicas de animales de laboratorio. Si

impacto es suficiente o no para invalidar la utilidad de dichas líneas, se puede determinar a partir de los resultados obtenidos en la presente tesis doctoral, aunque algunos estudios previos ya apuntan en sentido afirmativo (Casellas & Medrano, 2008; Casellas et al., 2014).

El efecto de la mayoría de las mutaciones solo se puede evaluar mediante métodos analíticos complejos y de forma indirecta, aprovechando la información contenida en los fenotipos (excepto para mutaciones de efecto grande). Dentro de este contexto, el objetivo principal de esta investigación fue ahondar en el desarrollo de una mejor comprensión relacionada con los patrones mutacionales en los mamíferos, con especial énfasis en la edad de los progenitores. En el caso de los ratones C57BL/6J, el modelo que presentó un mejor ajuste era el que incluía las varianzas mutacionales tanto paternas como maternas y además la varianza mutacional paterna evolucionaba con la edad de los padres. Esto implicaba dos fenómenos especialmente relevantes. En primer lugar el hecho que tanto los padres como las madres contribuían de manera independiente con nuevas mutaciones para la generación descendiente, y la carga mutacional procedente de cada sexo podía no ser la misma necesariamente, aunque este último punto no se pudo confirmar con los resultados obtenidos. En segundo lugar, la variabilidad mutacional se incrementó con la edad del padre, sugiriéndose que a mayor edad mayor acumulación de mutaciones, independientemente del signo y la magnitud de las mismas. Dado que no existen estudios previos de la misma índole en la literatura científica, resulta imposible discutir los resultados obtenidos más allá de la semejanza del conjunto de varianzas mutacionales con los valores reportados por otros autores previamente (Casellas y Medrano, 2008; Casellas et al., 2014; Nuzul et al., 2014), así como las revisiones llevadas a cabo por Lynch (1988) y Houle et al. (1996).

La varianza mutacional paterna se incrementó a medida que este envejecía, llegando hasta un incremento de ~45% durante el primer año de vida, mientras que no se observaron diferencias relevantes en el componente materno. Este fenómeno

no había sido descrito previamente en ratones, aunque si se había sugerido algún resultado compatible dentro del contexto de la herencia de algunas patologías con base genética. Hace ya más de un siglo, Weinberg (1912) estudió diferentes casos de acondroplasia y reportó que el último hijo nacido era más propenso a estar afectado. Este estudio sugería que la patología podía venir determinada componentes heredables, y que la edad de los progenitores tenía un papel principal en la aparición de la misma. Es la especie humana, donde existen estudios que pueden ser comparados con los realizados en la cepa consanguínea de ratones C57BL/6J, Goriely y Wilkie (2012), Forster et al. (2015) y Kong et al. (2012) han demostrado vínculos entre la edad paterna y la tasa de mutación (pero no entre la edad materna y la tasa de mutación).

Desde el punto de vista práctico la relevancia de las nuevas mutaciones así como el comportamiento de la varianza mutacional con relación al sexo y a la edad de los progenitores en la población de ratones de la línea C57BL/6J queda de manifiesto. Los resultados obtenidos en esta investigación amplían de una manera importante los obtenidos anteriormente (Casellas y Medrano, 2008; Casellas et al., 2014) y mejoran la comprensión de los mecanismos básicos que modulan la arquitectura genética de la especie. Considerando que fue creada como una línea isogénica, y que su utilización actual en la mayoría de estudios científicos se basa precisamente sobre esta premisa, los resultados obtenidos deben ser vistos como un toque de atención importante para la mayoría de laboratorios. Los ratones C57BL/6J no pueden considerarse homocigotos e isogénicos en absoluto, y su estabilidad genética a largo plazo no sería más que una quimera. Además, si se pretende minimizar la diversidad genética y el cambio genético en estas poblaciones, resultaría imprescindible minimizar la edad de los machos reproductores.

Resulta importante destacar que los resultados obtenidos en esta línea concreta de ratones debería ser directamente extrapolables a las demás especies de mamíferos implicados en procedimientos de laboratorio (e.g., cobayas, ratas,

hamsters, conejos, cerdos...). En la mayoría de casos, la homogeneidad, isogenicidad y estabilidad genética son características imprescindibles, que en ningún caso se podrían garantizar.

3. VARIANZA MUTACIONAL EN TERNOS DE RAZA BRUNA DE LOS PIRINEOS

La Bruna de los Pirineos es una raza vacuna de carne utilizada en sistemas de producción extensivos en los Pirineos catalanes. Para dar una idea de la importancia de esta población como fuente de datos para la presente tesis doctoral, debe considerarse que el censo aproximado es de unas 30.000 vacas de más de dos años, de las cuales 12.000 están inscritas en el Libro Genealógico y cerca de 5.500 participan del Programa de Control de Rendimientos (Fina, 2013). Desde el año 1990 hasta la actualidad se lleva a cabo un estricto programa de mejora genética con el propósito de mejorar los parámetros de fertilidad, habilidad materna, rusticidad en vacas, así como elevar los pesos al destete en terneros y los índices generales de crecimiento, además se procura obtener pesos promedios al nacimiento que no afecten la dificultad al momento del parto (Jordana y Pietrafita, 1993), de tal manera que la parametrización desarrollada en la presente investigación, así como los resultados obtenidos pueden ser un aporte de elevada connotación al sector ganadero productivo involucrado.

Para la consecución de los resultados, en este caso sobre el peso al nacimiento de los terneros de la raza Bruna de los Pirineos, se siguieron los mismos procedimientos analíticos que para la línea isogénica de ratones, obteniéndose unos patrones de resultados muy parecidos. Se validó la existencia de varianza mutacional tanto paterna como materna, así como la evolución lineal de la primera junto con la edad del toro, tal como pasaba con los ratones. Esto coincidiría nuevamente con los resultados reportados por Callaway (2012), Goriely y Wilkie (2012) y Kong et al. (2012) en humanos. De hecho es precisamente en la especie humana que está bien documentado que la edad paterna afecta de una manera importante a la integridad del ADN espermático. El envejecimiento trae como consecuencia la aparición de alteraciones de orden mutacional en la descendencia, con la subsiguiente presencia de alteraciones genéticas asociadas, y fracasos en el desarrollo embrionario

(Callaway, 2012; Goriely y Wilkie, 2012; Kong et al., 2012). En bovinos es poco lo que se ha investigado sobre el efecto de la edad paterna en relación con la acumulación de nuevas mutaciones, así como sobre los daños causados a nivel del ADN, y los defectos en el desarrollo embrionario. En un estudio desarrollado en bovinos *Bos indicus* de la raza Nellore, Carreira et al. (2017) reportaron una mayor susceptibilidad a los procesos oxidativos a nivel de ADN, así como un mayor grado de fragmentación espermática y mayor fragilidad del núcleo, a medida que los sementales de esta raza se envejecían. El nivel de condensación de la cromatina espermática en los toros de avanzada edad se ve aumentada con respecto a los toros más jóvenes, y esto parece ser contradictorio al presentar mayor susceptibilidad de sufrir daños oxidativos del ADN (Carreira et al., 2017). Sin embargo, se debe considerar la presencia de numerosos antioxidantes en los compartimientos post-testiculares (i.e., líquido epididimario, líquido seminal) que cumplen funciones de protección de las células espermáticas. Está bien documentado que a medida que avanza la edad en todos los mamíferos, los sistemas antioxidantes se vuelvan menos eficientes, permitiendo cada vez más frecuentemente la aparición de daños en el material genético.

Cabe resaltar ciertos aspectos de interés que dan una connotación importante a los resultados obtenidos. La implicación práctica que deriva de este estudio para la especie bovina, se relaciona con la selección de sementales y el efecto que puede tener el uso de animales de edad avanzada sobre el valor genético de las futuras generaciones. Dado que la variabilidad mutacional paterna se incrementaría con la edad del semental, la probabilidad de obtener animales extremos en términos de valor genético se incrementaría con los sementales más viejos. Esto contradiría la práctica habitual, dónde la edad de los toros no suele representar un criterio relevante a la hora de escoger los reproductores. En general, la vida útil de un toro reproductor casi siempre está determinada por las condiciones físicas del mismo, así como por la presencia de hijas en uso en los lotes de servicio. Paralelamente, la inclusión de los efectos mutacionales en los modelos de evaluación genética permiten una predicción

más precisa del componente genético de los animales, tal como se ha descrito anteriormente en ovinos (Casellas et al., 2010), pudiendo tener esto consecuencias importantes sobre el progreso genético de las razas ganaderas.

4. IMPLICACIONES DE LA VARIANZA MUTACIONAL

La relevancia de las nuevas mutaciones en relación con el origen genético de los mamíferos ha sido señalada repetidamente a lo largo de las últimas décadas (Hill 1982a,b; Caballero et al. 1991; Keightley y Hill, 1992; Casellas y Medrano 2008). Los resultados obtenidos en este trabajo de investigación corroboran que la varianza genética aditiva, en las poblaciones estudiadas, era relativamente alta, tanto en términos de la varianza genética aditiva ya existente en la generación fundadora, como en el caso de la varianza originada de las nuevas mutaciones. Mientras que la varianza genética aditiva ocupó un porcentaje importante de la varianza total, la varianza mutacional tanto paterna como materna presentaba valores mucho más bajos, aunque sus intervalos de confianza descartaban el valor nulo y confirmaban su relevancia a nivel biológico. Por otra parte, la varianza mutacional paterna se sugirió más elevada que la materna, aunque sus intervalos de confianza se solaparon. Este hecho previene de formular otras conclusiones sobre el predominio de cualquier sexo desde el punto de vista de la contribución mutacional a la descendencia.

Aun así, las diferencias entre las líneas germinales de la madre y del padre fueron evidenciadas cuando se compararon los efectos de la edad en la tasa de nuevas mutaciones. La carga mutacional paterna se incrementó a medida que este envejecía, fenómeno que no se observó en el sexo femenino. Aunque, el coeficiente de regresión relacionado con la edad del semental puede parecer bajo, su intervalo de confianza evidentemente descartó el valor nulo e implicó incrementos importantes a lo largo del tiempo en ambas especies de mamíferos. Esto se ha sugerido previamente en otras especies como la humana, dentro del ámbito de la base genética del autismo (Callaway, 2012), la esquizofrenia, el síndrome de Apert, y algunos carcinomas (Arroyo et al., 1999; Bertram, 2000; Goriely y Wilkie, 2012; Kong et al., 2012); paralelamente, para la contribución mutacional de la madre, tampoco se observaron cambios ligados con su edad.

Otros estudios como los reportados por Mahera et al. (2016), sugirieron también que cambios puntuales *de novo* en el ADN, no heredados de los progenitores, suelen producirse en la línea germinal masculina y su frecuencia aumenta con la edad. La explicación de tal fenómeno radica en que los espermatozoides se producen a partir de las espermatogonias. Este es un tipo celular “madre”, que al dividirse genera una copia de sí misma y una célula que se diferenciará, dando lugar a los espermatozoides. Por una parte se producen nuevas células “madre” que persisten en la base del ciclo y sufren mitosis sucesivas, mientras que por otra parte se originan células destinadas a diferenciarse y, en consecuencia, a sufrir una meiosis completa. Dado el volumen de producción de células germinales en los machos, las espermatogonias se dividen a menudo. En cada división se debe copiar todo el material hereditario, con la posibilidad de que se produzca un error, una mutación que puede ser inocua o con capacidad de modificar de alguna manera algún carácter fenotípico del individuo, llegando incluso a originar enfermedades. Este argumento representaría la base biológica y fisiológica que sustentaría la evolución con la edad de la tasa de mutación en los sementales.

El aporte de nuevas mutaciones en función de la edad para el caso de las hembras no mostró ningún tipo de incremento ni disminución: Para este caso, el aporte mutacional se podía considerar invariable a medida que las hembras envejecieron, tal como han sugerido antes otros autores (Murray et al., 2012). De hecho, se señala que este comportamiento obedece al hecho de que el proceso de la oogénesis inicia la meiosis durante el desarrollo fetal y todos los oocitos primarios quedan parados en la profase I ya antes del nacimiento. No será hasta que la hembra alcance la pubertad que uno o varios oocitos retomaran los procesos de la meiosis hasta finalizarla en cada ciclo reproductivo, sin requerir de divisiones adicionales en estas células. De esta manera, la probabilidad de acumular mutaciones es claramente inferior que en caso de los espermatozoides, al implicar un número significativamente menor de mitosis.

Los resultados obtenidos en el presente estudio se pueden extrapolar a la especie humana, corroborando las conclusiones obtenidas previamente por Arroyo et al. (1999), Callaway (2012), Goriely y Wilkie (2012), Kong et al. (2012) y Forster et al. (2015). En caso de aplicarse la parametrización desarrollada en la presente tesis doctoral a nuestra especie, se podría recurrir a los registros recolectados en maternidades y centros de salud, que guarden la información de madres y padres, así como de la descendencia referentes ciertas característica físicas y presencia de algunas posibles alteraciones de tipo genéticos atribuibles a procesos mutacionales.

BIBLIOGRAFÍA

- Arroyo Carrera ML, Martínez-Frías JJ, Marco Pérez L, Paisán-Grisolía A, Cárdenes Rodríguez C, Nieto Conde V, Félix Rodríguez JJ, Egüés Jimeno MC, Morales Fernández J, Gómez-Ullate Vergara M, Pardo Romero A, Peñas Valiente MJ, Oliván del Cacho A, Lara Palma L. 1999. Síndrome de Apert: análisis clínico-epidemiológico de una serie consecutiva de casos en España. *Medicina Fetal y Neonatología* **51**: 67-672.
- Bertram J. 2000. The molecular biology of cancer. *Mol Aspects Med* **21**: 167-223.
- Blasco A. 2001. The Bayesian controversy in animal breeding. *J Anim Sci* **79**: 2023-2046.
- Caballero A, Toro MA, López-Fanjul C. 1991. The response to artificial selection from new mutations in *Drosophila melanogaster*. *Genetics* **127**: 89-102.
- Callaway E. 2012. Genome study may explain links between paternal age and conditions such as autism. *Nature* **488**: 439.
- Carreira J, Trevizan I, Carvalho B, Kipper L, Rodriguez C, Silva S, Perri J, Drevet J and Koivisto M. 2017. Does sperm quality and DNA integrity differ in cryopreserved semen samples from young, adult, and aged Nellore bulls. *Basic Clin Androl* **27**: 12.
- Casellas J, Caja G, Piedrafita J. 2010. Accounting for additive genetic mutations on litter size in Ripollesa sheep. *J Anim Sci* **88**: 1248-1255.
- Casellas J, Medrano JF. 2008. Within-generation mutation variance for litter size in inbred mice. *Genetics* **179**: 2147-2155.
- Casellas J, Gianola D, Medrano JF. 2014. Bayesian analysis of additive epistasis

- arising from new mutations in mice. *Genet Res* **96**: e008.
- Festing MFW. 1996. Origins and characteristics of inbred strains of mice. Páginas 1537-1576 en *Genetic variants and strains of the laboratory mouse* (editores. Lyon MF, Rastan S, Brown SDM). Oxford University Press, Oxford, Reino Unido.
- Festing MFW. 1998. *Inbred strains of mice and their characteristics*. Mouse Genome Informatics. The Jackson Laboratory, Bar Harbor, ME, EUA.
- Fina M. 2013. *Optimització de l'avaluació genètica de la raça bovina Bruna dels Pirineus*. Tesis Doctoral, Universitat Autònoma de Barcelona, Bellaterra, Barcelona. España.
- Forster P, Hohoff C, Dunkelmann B, Schürenkamp M, Pfeiffer H, Neuhuber F, Brinkmann B. 2015. Elevated germline mutation rate in teenage fathers. *Proc R Soc B* **282**: 1-7.
- Gelfand A, Smith AFM. 1990. Sampling based approaches to calculating marginal densities. *J Am Stat Assoc* **85**: 398-409.
- Goriely A, Wilkie AO. 2012. Paternal age effect mutations and selfish spermatogonial selection: causes and consequences for human disease. *Am J Hum Genet* **90**: 175-200.
- Hastings WK. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**: 97-109.
- Henderson CR. 1973. Sire evaluations and genetic trends. Páginas 10-41 en *Proc Anim Breed Genet Symp in Honor of Jay L Lush*. American Society of Animal Science and American Dairy Science Association, Champaign, Illinois, EUA.

- Hill WG. 1982a. Rates of change in quantitative traits from fixation of new mutations. *Proc Natl Acad Sci USA* **79**: 142-145.
- Hill, WG. 1982b. Predictions of response to artificial selection from new mutations. *Genet Res* **40**: 255-278.
- Houle D, Morikawa B, Lynch M. 1996. Comparing mutational variabilities. *Genetics* **143**: 1467-1483.
- Jeffreys H. 1961. *Theory of Probability*. Clarendon Press, Oxford, Reino Unido.
- Kass RE, Raftery AE. 1995. Bayes factors. *J Am Stat Assoc* **90**: 773-795.
- Keightley PD, Hill WG. 1992. Quantitative genetic variation in body size of mice from new mutations. *Genetics* **131**: 693-700.
- Kong A, Frigge M, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson S, Sigurdsson A, Jonasdottir A. 2012. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**: 429-550.
- Lynch M. 1988. The rate of polygenic mutation. *Genet Res* **51**: 137-148.
- Lynch M, Conery J, Bürger R. 1995. Mutation accumulation and the extinction of small populations. *Am Nat* **146**: 489-518.
- Mahera G, McGowanb S, Giannoulatoua E, Verrillc C, Goriely A, Wilkie A. 2016. Visualizing the origins of selfish de novo mutations in individual seminiferous tubules of human testes. *Proc Natl Acad Sci USA* **113**: 2454 –2459.
- Metropolis R, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953. Equations of state calculations by fast computing machines. *J Chem Phys* **21**: 1087-1092.
- Murray RK, Rodwell VW, Bender D, Botham KM, Weil PA, Kennelly PJ. 2012.

Harper's Illustrated Biochemistry. McGraw Hill Professional, New York, NY, EUA.

Newton MA, Raftery AE. 1994. Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *J Royal Stat Soc B* **56**: 1-48.

Nuzul W, Nielsen VH, Jensen J. 2014. Estimation of genetic variance components including mutation and epistasis using Bayesian approach in a selection experiment on body weight in mice. Páginas 488-491 en *Proc 10th World Congr Genet Appl Anim Prod*, Vancouver, Canadá.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. 2002. Bayesian measures of model complexity and fit. *J Royal Stat Soc B* **64**: 583-639.

Weinberg W. 1912. Zur Vererbung des Zwergwuchses. *Arch. Rassen Gesellschafts Biol.* **9**: 710-718.

Wray NR. 1990. Accounting for mutation effects in the additive genetic variance-covariance matrix and its inverse. *Biometrics* **46**: 177-186.

Wright S. 1922. Coefficients of inbreeding and relationship. *Am Nat* **56**: 330-338.

CAPÍTULO VI
CONCLUSIONES

Una vez expuestos los resultados de las investigaciones que formaron parte integral de la presente tesis doctoral y habiendo realizado las discusiones pertinentes, a continuación se presentan las principales conclusiones obtenidas, las cuales se agrupan de acuerdo a los ítemes de relevancia desarrollados a lo largo de este trabajo investigador tomando en consideración la metodología analítica y las poblaciones evaluadas.

PARAMETRIZACIÓN DEL MODELO ANALÍTICO DESARROLLADO

- 1) El modelo desarrollado representa una contribución importante en el campo de la genética, permitiendo abordar de manera novedosa la varianza mutacional aditiva; permite la inclusión de términos mutacionales específicos para cada progenitor e incorpora un segundo nivel de jerarquización que considera la posibilidad de cambios en la variabilidad mutacional en función de la edad del reproductor.
- 2) Los requerimientos del análisis se centran únicamente en la disponibilidad de registros fenotípicos y genealógicos, siendo un modelo altamente flexible y versátil, y pudiéndose implementar junto con procedimientos estándar de comparación de modelos.

CEPA C57BL/6J

- 3) Tanto los padres como las madres contribuyeron nuevas mutaciones con efectos relevantes sobre el peso al destete de la generación de descendientes. Su magnitud era pequeña pero estadísticamente relevante, y no se podía establecer una preeminencia clara para alguno de los dos sexos, aunque se sugería una mayor variabilidad mutacional procedente de los machos.

4) La variabilidad mutacional procedente del padre se incrementaba de manera lineal con la edad del mismo, originándose incrementos cercanos al 50% durante el primer año de vida.

5) Deberían revisarse los protocolos de mantenimiento y gestión de las cepas de isogénicas de animales de laboratorio con el objetivo de minimizar la variabilidad genética y el cambio genético. A tal efecto se sugiere la necesidad de utilizar machos jóvenes.

RAZA BOVINA BRUNA DE LOS PIRINEOS

6) La variabilidad mutacional con efecto sobre el peso al nacimiento de los terneros procedía tanto del semental como de la vaca, sin diferencias relevantes entre ambas fuentes de variación.

7) Nuevamente la variabilidad mutacional con origen paterno se incrementaba con la edad del macho reproductor, mientras que la de origen materno no sugería ningún tipo de cambio a lo largo de la vida reproductiva de la vaca.

8) Con el objetivo de maximizar la probabilidad de generar animales extremos en términos de mérito genético, debería considerarse la posibilidad de promocionar el uso de sementales de avanzada edad o, en caso de ser posible, almacenar dosis seminales para su uso mediante inseminación artificial, antes de descartar los sementales por criterios de edad.

