ISSN: 1130-6009

ITEA

PRODUCCIÓN ANIMAL

Asociación Interprofesional para el Desarrollo Agrario

2000 – Vol. 96A N.º 3



I T E A

Información Técnica Económica Agraria
Revista de la Asociación Interprofesional para el Desarrollo Agrario

2000 - AÑO XXXI Vol. 96A N.º 3	**DIRECCIÓN Y REDACCIÓN** Montañana, 176 - Apartado 727 50080 ZARAGOZA (ESPAÑA) Tel.: 34-976 576311 Fax.: 34-976 575501 E-mail: enotivol@aragob.es andreu@eead.csic.es cmmarin@aragob.es	Depósito legal: Z-577-82 ISSN: 1130-6009 INO Reproducciones, S.A. Polígono Miguel Servet, nave 13 50013 Zaragoza

COMITÉ DE REDACCIÓN

DIRECTOR: Eduardo Notivol Paíno
SUBDIRECTOR: Joaquín Uriarte Abad
SECRETARIAS: Serie Producción Vegetal: Pilar Andreu Puyal
Serie Producción Animal: Clara Marín Alcalá
VOCALES: José Álvarez Álvarez
Rafael Delfa Belenguer
Joaquín Gómez Aparisi
Emilio Manrique Persiva
José Luis Alabart Álvarez
Juan A. Marín Velázquez
Azucena Gracia Royo
M.ª Dolores Quílez Sáez de Viteri
Carlos Zaragoza Larios
Pere Alberti Lasalle

JUNTA DIRECTIVA DE A.I.D.A.

PRESIDENTE: Leonardo Plana Claver
VICEPRESIDENTES: 1.º Emilio Manrique Persiva
2.º Rafael Socias i Company
SECRETARIO: José Álvarez Álvarez
TESORERO: Joaquín Uriarte Abad
VOCALES: José Folch Pera
Ignacio Delgado Enguita
Ricardo Revilla Delgado
Joaquín Gómez Aparisi
Antonio Felipe Mansergas
Pere Alberti Lasalle
Dunixi Gabiña Iturriaga

Prohibida toda reproducción total o parcial sin autorización expresa
de la Asociación Interprofesional para el Desarrollo Agrario, editor titular del Copyright.

ITEA no se responsabiliza necesariamente con las opiniones vertidas en los artículos firmados
que publica, cuya responsabilidad corresponde a sus autores.

Suscripciones y Distribución

Información Técnica Económica Agraria publica tres números en volumen. En 2000 se publicarán los volúmenes 96A y 96V correspondientes a las series Producción Animal y Producción Vegetal.

El precio de la suscripción para 2000 será de 4.000 ptas. ó 24€ para una serie y de 5.500 ptas. ó 33€ para las dos series.

Se acepta el intercambio con otras revistas.

ITEA. Apartado 727. 50080 Zaragoza (ESPAÑA)

PRESENTACIÓN

En la segunda semana del mes de junio de 2000 tuvo lugar en Caldes de Montbui (Barcelona) la X Reunión Nacional de Mejora Genética Animal, organizada por la Unidad de Cunicultura del I. R. T. A. (Instituto de Investigación y Tecnología Agroalimentarias de Cataluña) y el Departamento de Ciencia Animal y de los Alimentos de la Universidad Autónoma de Barcelona. Las sesiones se desarrollaron en el marco de la Granja Escuela Torre Marimón de la Diputación de Barcelona.

La idea inicial de esta convocatoria era reunir el mayor número de profesionales relacionados con la mejora genética animal (centros de investigación, universidades y empresas), creando las condiciones para el intercambio de información y, más aún, la discusión de los trabajos en ejecución o en proyecto. Las Jornadas se estructuraron a partir de las comunicaciones orales (4 sesiones) y ponencias invitadas. Además de ser un lugar de encuentro, se intentó que permitiera al mismo tiempo un cierto grado de "formación", con dos características: pluridisciplinar y de vanguardia.

Si bien es cierto que el concepto de gen ha generado una separación aparente entre Genética Molecular (estudio estructura genes) y Genética Cuantitativa (estudio variación poligénica), se ha tornado en colaboración necesaria entre ellas para el estudio de los mapas genéticos y la búsqueda de QTL's. La ponencia del Dr. Arranz (Universidad de León, España) para la detección de regiones cromosómicas candidatas con *genome scan*, y la del Dr. Rothschild (Universidad de Iowa, EEUU, Coordinador del genoma porcino) con aplicación al estudio de genes candidatos en porcino permitieron una diferente aproximación a la búsqueda de genes de interés.

Las ponencias más "cuantitativas" fueron desarrolladas por el Dr. Ducrocq (INRA de Jouy-en-Josas, Francia) y el Dr. Gianola (Universidad de Wisconsin-Madison, EEUU). El primero incidiendo sobre la dificultad que entraña el análisis de los datos de supervivencia y presentando un programa de aplicación, y el segundo proponiendo un método para poder realizar inferencias sobre la varianza genética aditiva (u otros parámetros) en el transcurso de los experimentos de selección desde un enfoque bayesiano.

A lo largo de las sesiones se abordaron temas relacionados con estudios genéticos en las diferentes especies de interés ganadero, tanto a nivel molecular, como de caracterización de

las poblaciones sometidas a selección o utilizadas en programas de cruzamientos, como a nivel de desarrollos teóricos.

Aprovechamos aquí la oportunidad que I. T. E. A. nos brinda para la publicación del contenido de las jornadas con objeto de facilitar la difusión hacia los colegas de otros campos de investigación relacionados con la producción animal o con la mejora genética.

Por último, agradecer la colaboración de la Diputación de Barcelona, así como la financiación proporcionada por el I. R. T. A., y las ayudas de la C. I. R. I. T. de la Generalitat de Catalunya y del Ministerio de Ciencia y Tecnología.

Ernesto A. Gómez Blasco
Investigador Responsable X Jornadas Nacionales de Mejora Genética Animal

ÍNDICE

Página

M. AMILLS, N. JIMÉNEZ, A. RICCARDI, J. FOLCH, A. SÁNCHEZ.- Clonación y secuenciación de alelos MHC-DRB de cabra montesa (*capra pyrenaica*) 169

J.J. ARRANZ.- Utilización de *genome scan* para la detección de regiones cromosómicas que expliquen parte de la varianza aditiva y el posterior estudio de genes candidatos 173

J.A. BARO DE LA FUENTE, R. ÁLVAREZ, C.E. CARLEOS, D. GARCÍA, H. LAMELAS.- RePed. A tool for checking, exploring and debugging pedigrees 184

J.A. BARO, C.E. CARLEOS, R. PONG-WONG, R. ÁLVAREZ.- Evaluación bajo la hipótesis de herencia mixta; uso de información de QTLs en la predicción de valores genéticos poligénicos ... 189

A. BARROSO, L.J. ROYO, J. CAÑÓN, S. DUNNER.- La hipertrofia muscular hereditaria: génesis de alelos dominantes negativos de la miostatina (GDF-8) Murina 191

A. BLASCO.- La perspectiva bayesiana. Tres nuevos diálogos entre Hylas y Filonús 199

A. CLOP, G. NYMAN, A. SÁNCHEZ, J. L. NOGUERA, L. ANDERSSON.- Detección del origen filogenético de diferentes poblaciones porcinas españolas por pirosecuenciación de una región del citocromo B del ADN mitocondrial porcino 220

M.L. CHECA, C. CARLEOS, J.A. BARO, S. DUNNER, J. CAÑÓN.- Estimación de frecuencias alélicas en *pools* de ADN mediante el procedimiento de señales de electroforesis capilar 225

V. DUCROCQ, E.A. GÓMEZ.- Análisis de supervivencia, una herramienta estadística para datos de longevidad 235

A. FERNÁNDEZ, C. ÓVILO, C. RODRÍGUEZ, L. SILIÓ.- Polimorfismo *Hha*I del gen de la hormona de crecimiento en cerdos ibéricos: relación con caracteres productivos 254

A. FERNÁNDEZ, M.A. TORO, J. RODRIGÁÑEZ, M.C. RODRÍGUEZ, L. SILIÓ.- La fuerza del mestizo: estimación de la heterosis en el crecimiento de cerdos ibéricos en Montanera 261

M.T. FERNÁNDEZ, R. ALENDA, M. SERRANO, E. UGARTE.- Estudio del carácter velocidad de ordeño y recuento de células somáticas en primera lactación de la población vacuna frisona del País Vasco 270

D. GARCÍA, C. CARLEOS, J.A. BARO, J. CAÑÓN.- Identificación de alelos con efecto significativo en la diferenciación de poblaciones 277

D. GARCÍA, M.L. CHECA, P. GARCÍA-ATANCE, S. DUNNER, J. CAÑÓN.- Medidas de diversidad genética en poblaciones de caballos celtas españoles 282

L.A. GARCÍA CORTÉS, C. CABRILLO.- Sobre la convergencia del muestreo de Gibbs en los modelos bayesianos jerárquicos 290

E.A. GÓMEZ, O. RAFEL, J. RAMÓN, O. PERUCHO.- Propuesta de estimación de parámetros genéticos del carácter índice de conversión en dos líneas de conejos seleccionadas por crecimiento 296

M. HABIB YAHYAOUI, A. COLL, A. SÁNCHEZ, J.M. FOLCH.- Caracterización de variantes genéticas en la κ-caseína caprina 302

Página

A. LEGARRA, E. UGARTE, E. RUIZ DE ZÁRATE.- Adecuación de la estructura de datos del caballo de carne en Álava para una estima de componentes de varianza 311

P. LÓPEZ-ROMERO, M.J. CARABAÑO.- Evaluación de modelos alternativos para el análisis de datos de producción de leche a lo largo de la lactación 318

M. PILES, D. GIANOLA, L. VARONA, A. BLASCO.- Análisis de la respuesta correlacionada a la selección por velocidad de crecimiento en las curvas de crecimiento de conejos 337

M.F. ROTHSCHILD.- Advances in pig molecular genetics, gene mapping and genomics 349

M. SERRANO, P. DÍEZ DE TEJADA, D. HERNÁNDEZ, J.J. JURADO.- El esquema de selección de la raza caprina "cabra del Guadarrama" 362

M.A. TORO, C. BARRAGÁN, J. RODRIGÁÑEZ, C. RODRÍGUEZ, L. SILIÓ.- Estimación del parentesco en cerdos ibéricos utilizando marcadores moleculares 370

L. VARONA, G. DAVALOS, M. PÉREZ-ENCISO, J.M. FOLCH, N. JIMÉNEZ, A. SÁNCHEZ, J.L. NOGUERA.- Análisis bayesiano multivariante de componentes de varianza de caracteres de producción en una población Pietrain usando marcadores moleculares 379

G. YAGÜE-UTRILLA, C. MORENO, L.A. GARCÍA-CORTÉS, J. ALTARRIBA.- Inferencia en poblaciones donde se desconoce la cantidad de información perdida 383

ITEA (2000). Vol. **96A N.° 3,** 169-172

CLONACIÓN Y SECUENCIACIÓN DE ALELOS MHC-DRB DE CABRA MONTESA (*CAPRA PYRENAICA*)

M. Amills*, N. Jiménez*, A. Riccardi*, J. Folch, A. Sánchez***

* Unitat de Genètica, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra 08193, España

** Unidad de Tecnología en Producción Animal, Servicio de Investigación Agroalimentaria, Diputación general de Aragón, Apdo. 727, 50080 Zaragoza, España

RESUMEN

En el presente trabajo, se ha caracterizado el polimorfismo del complejo mayor de histocompatibilidad (MHC) en la cabra montesa *(Capra pyrenaica)*. Con esta finalidad, se secuenció el exón 2 del gen *DRB* en seis individuos no emparentados procedentes de distintas áreas geográficas. Únicamente se detectaron 2 tipos alélicos que presentaban una similitud nucleotídica del 83%. Ello indica que el polimorfismo del gen *DRB* en la cabra montesa es muy limitado, probablemente debido a los cuellos de botella poblacionales que esta especie sufrió durante los s. XIX y XX. Estos resultados serán de gran interés en la caracterización genética de la cabra montesa, contribuyendo a establecer un criterio de selección, respecto a la reintroducción de esta especie en áreas geográficas en las que su tamaño efectivo se haya reducido significativamente, basado en la utilización de marcadores genéticos.

SUMMARY

We have characterized the polymorphism of the major histocompatibility complex (MHC) in the Spanish ibex *(Capra pyrenaica)* by sequencing the second exon of the *DRB* gene in six unrelated individuals from different geographic areas. Our results demonstrate that the polymorphism of the *DRB* gene is very reduced in the Spanish ibex, with only two major allelic lineages which show 83% nucleotide similarity. These results are likely explained by the severe population bottlenecks that this species suffered in the XIX and XX centuries. These results will contribute to the genetic characterization of the Spanish ibex and to the establishment of selection criteria for the reintroduction of this species.

Introducción

La cabra montesa *(Capra pyrenaica)* es un cáprido salvaje autóctono de la Península Ibérica estrechamente emparentado con el ibex alpino *(Capra ibex)*. La distribución geográfica de esta especie comprende los macizos montañosos del centro (Gredos y Batuecas) y este (Beceite-Tortosa, Muela de Cortés, Cazorla, Nevada y Ronda) de la Península. Aunque inicialmente existían cuatro subespecies de cabra montesa, *C. p. hispanica* (CPH), *C. p. victoriae* (CPV), *C. p. lusitanica* (CPL) y *C. p. pyrenaica* (CPP), la caza excesiva y la progresiva destrucción de su hábitat natural posiblemente propició la extinción de CPL y CPP (MANCEAU, 1992).

La caracterización genética de la cabra montesa constituiría una herramienta fundamental para diseñar estrategias óptimas de repoblación en aquellas zonas geográficas en las que esta especie haya experimentado un declive demográfico. De este modo, podría implementarse la reintroducción de nuevos ejemplares que aporten variabilidad genética adicional a las poblaciones receptoras, incrementando su eficacia biológica global. El gen *DRB* del complejo mayor de histocompatibilidad (MHC) tiene un profundo efecto sobre la resistencia genética a numerosas enfermedades y además presentan niveles muy elevados de polimorfismo, por lo que constituiría un excelente marcador genético para seleccionar ejemplares destinados a la repoblación de áreas que posean un bajo número efectivo de individuos. En el presente trabajo, se ha evaluado la variabilidad del exón 2 del gen MHC-*DRB* de cabra montesa (*Capy-DRB*) mediante clonación y secuenciación de producto *DRB* amplificado a partir de seis individuos de distinta procedencia geográfica y subespecie.

Material y métodos

Los ejemplares de cabra montesa analizados, su procedencia geográfica y subespecie, y el alelo *DRB* identificado en cada uno de ellos pueden observarse en el cuadro 1. La amplificación del segundo exón del gen DRB se realizó de acuerdo a Amills et al (1996). El producto amplificado fue clonado en un plásmido pCR2.1-TOPO (Invitrogen) y secuenciado en ambas direcciones mediante el kit ABI PRISM Cycle Sequencing kit (Perkin Elmer Biosystems). Las reacciones de secuenciación fueron analizadas en un secuenciador ABI 373 (Perkin Elmer Biosystems). Los alineamientos de las secuencias se realizaron mediante el programa CLUSTALW.

Resultados y discusión

La molécula DR del MHC de clase II tiene como función primordial la presentación de péptidos antigénicos a los linfocitos T *helper*, jugando por ello un papel clave en el desarrollo de la respuesta inmune adaptativa (Hughes y Yeager 1998). Los dominios α_1 y β_1 de la molécula DR codifican el lugar de unión al antígeno o ABS (*antigen-binding site*), una estructura en forma de surco capaz de acomodar péptidos de 13-21 aminoácidos para que puedan ser reconocidos por el receptor del linfocito T (HUGHES y YEAGER, 1998). El exón 2 del gen *DRB*

Cuadro 1
Procedencia, subespecie y alelo identificado en cada uno de los individuos analizados

Individuo	Procedencia	Subespecie	Alelo
128	Muela de Cortés	CPH	Capy-DRB1.1
B24	Beceite-Tortosa	CPH	Capy-DRB1.2
187	Batuecas	CPV	Capy-DRB1.3
66	Gredos	CPV	Capy-DRB1.4
150	Beceite-Tortosa	CPH	Capy-DRB1.5
183	Batuecas	CPV	Capy-DRB1.6

codifica el dominio β_1 de la molécula DR y posee una elevada variabilidad en especies como la vaca (63 alelos), la oveja (70 alelos) y la cabra doméstica (28 alelos) (AMILLS *et al.*, 1998). Paradójicamente, los resultados obtenidos en cabra montesa demuestran que los niveles de variabilidad del exón 2 del gen *DRB* son muy bajos en esta especie (ver cuadro 1). Sólo se encontraron 2 tipos mayoritarios de alelos: uno de ellos incluye los alelos *Capy-DRB1.1* y *Capy-DRB1.2*, mientras que el otro comprende los alelos *Capy-DRB1.3*, *Capy-DRB1.4*, *Capy-DRB1.5* y *Capy-DRB1.6*. La similitud nucleotídica promedio entre ambos grupos de aproximadamente el 83%. La similitud nucleotídica promedio intragrupo, en cambio, fue muy elevada y oscilaba entre el 98-99,5%.

Los resultados obtenidos indican bajos niveles de polimorfismo en el gen *DRB* y concuerdan con los datos obtenidos por JIMÉNEZ *et al.* (1999), que indicaban la existencia de bajos niveles de variabilidad genética en las distintas subespecies de cabra montesa. Las poblaciones de cabra montesa han sufrido importantes cuellos de botella poblacionales durante los s. XIX y XX, con tamaños efectivos que en algunos núcleos no sobrepasaban los 30 individuos (MANCEAU, 1992). Probablemente, ello haya favorecido el mantenimiento de determinados tipos alélicos por un mero efecto de deriva genética.

En general, se asume que el extraordinario polimorfismo detectado en los genes MHC de humano, ratón y otras especies resulta necesario para asegurar la presentación efectiva de un amplio espectro de péptidos de distinta secuencia y longitud a los linfocitos T (HUGHES y YEAGER, 1998). En otras palabras, la gran variabilidad genética del MHC sería una condición necesaria para que estas especies animales pudieran generar una respuesta inmune adaptativa eficiente contra una gran variedad de microorganismos patógenos (MIKKO *et al.*, 1999). No obstante, existen numerosos ejemplos de especies animales (*Alces alces, Castor fiber, Panthera leo persica, Ovibos moschatus* etc.) en las que a pesar de que existe un reducido polimorfismo a nivel del MHC (MIKKO *et al,*. 1999) no se observa una mayor susceptibilidad a contraer enfermedades infecciosas. Aunque la resolución de esta aparente paradoja todavía sigue pendiente, la caracterización del gen *DRB* y otros genes MHC en la cabra montesa contribuirá decisivamente a comprender mejor la interacción de esta especie con los microorganismos patógenos de su entorno natural. Asimismo, el gen *DRB* podría ser incluido como marcador genético para seleccionar individuos destinados a repoblar áreas geográficas en las que esta especie escasee.

Bibliografía

AMILLS *et al.*, 1996. Vet. Immunol. Immunopathol. 55: 255-260

AMILLS *et al.*, 1998. Rev. Sci. Tech. 17:108-20.

HUGHES y YEAGER, 1998. Annu. Rev. Genet. 32: 415-435.

JIMÉNEZ *et al.*, 1999. ITEA 20: 300-302.

MANCEAU, 1992. Polymorphisme des sequences d'ADN mitochondrial dans le genre Capra. Application a la conservation du bouquetin des Pyréneées. Tesis Doctoral (CNRS UMR 5553).

MIKKO *et al.*, 1999. Immunol. Rev. 167: 169-178.

```
                1                                                  50
Capy-DRB1.1 CTGGAGTATT ATAAGGCCGA GTGTCATTTC TCCAACGGGA CCGGGCGGGT
Capy-DRB1.2 ---------- ---------- ---------- ---------- ----------
Capy-DRB1.3 ---AG----- G----AGA-- ---------- -----T---- ---A------
Capy-DRB1.4 ---AG----- G----AGA-- ---------- -----T---- ---AA-----
Capy-DRB1.5 ---AG----- G----AGA-- ---------- -----T---- ---A------
Capy-DRB1.6 ---AG----- G----AGA-- ---------- -----T---- ---A------

               51                                                 100
Capy-DRB1.1 GCGGTTGCTG CACAGATTCT ACACTAATGG AGAAGAGACC GTGCGCTTCG
Capy-DRB1.2 ---------- ---------- ----C----- ---------- ----------
Capy-DRB1.3 ----C-T--- G------A-- T-CAT----- G-------T- ----------
Capy-DRB1.4 ----C-T--- G------A-- T-CAT----- --------T- ----------
Capy-DRB1.5 ----C-T--- G------A-- T-CAT----- --------T- ----------
Capy-DRB1.6 ----C-T--- G------A-- T-CAT----- --------T- ----------

              101                                                 150
Capy-DRB1.1 ACAGCGACTG GGGCGAGTTC CGGGCGGTGA CCGAGCAGGG GCAGGAGGAC
Capy-DRB1.2 ---------- ---------- ---------- ---------- ----------
Capy-DRB1.3 ---------- ---------- -----A---- -------T--- --G-CC---T
Capy-DRB1.4 ---------- ---------- -----A---- -------T--- --G-CC---T
Capy-DRB1.5 --G------- ---------- -----A---- -------T--- --G-CC---T
Capy-DRB1.6 ---------- ---------- -----A---- -------T--- --G-CC---T

              151                                                 200
Capy-DRB1.1 GCCGAGTACT GGAACAGCCA GAAGGACTTC CTGGAGCAGA AGCGGGCCGA
Capy-DRB1.2 ---------- ---------- ---------- ---------- ----------
Capy-DRB1.3 ---------- ---------- ------GA-- ------AGC- G-A--A---C
Capy-DRB1.4 ---------- ---------- ------GA-- ------AGC- G-A--A---C
Capy-DRB1.5 ---------- ---------- ------GA-- ------AGC- G-A--A---C
Capy-DRB1.6 ---------- ---------- ------GA-- ------AGC- G-A--A---C

              201                                  237
Capy-DRB1.1 GGTGGACACG GTGTGCAGAC ATAACTACGG GGTCCTT
Capy-DRB1.2 ---------- ---------- ---------- -------
Capy-DRB1.3 ---------- TAC------T -C--------- ----GG-
Capy-DRB1.4 ---------- TAC------T -C--------- ----GG-
Capy-DRB1.5 ---------- TAC------T -C--------- ----GG-
Capy-DRB1.6 ---------- TAC------T -C--------- ----GG-
```

Figura 1. Alineamiento CLUSTALW de secuencias *Capy-DRB* (exón 2)

ITEA (2000), Vol. 96A N.° 3, 173-183

UTILIZACIÓN DEL *GENOME SCAN* PARA LA DETECCIÓN DE REGIONES CROMOSÓMICAS QUE EXPLIQUEN PARTE DE LA VARIANZA ADITIVA Y EL POSTERIOR ESTUDIO DE GENES CANDIDATOS

J.J. Arranz

Dpto. Producción Animal I, Facultad de Veterinaria, Universidad de León, 24071 León, España

Introducción

Los caracteres productivos objeto de selección en las especies domésticas son, en general, de tipo cuantitativo. La variación observada en el fenotipo es el resultado de la acción directa de un número grande de genes con efecto individual pequeño y aditivo, de la interacción entre alelos del mismo o distinto gen, del efecto de factores ambientales diversos y de la interacción entre genes y ambiente (FALCONER y MACKAY, 1996). La contribución a la varianza aditiva de estos genes, denominados genéricamente QTLs (iniciales del término inglés *Quantitative Trait Loci*), ha sido tema de estudio desde los inicios de la investigación genética (SAX, 1923). El conocimiento de la arquitectura molecular de estos caracteres productivos, podría incrementar la respuesta a la selección y como consecuencia su eficacia, disminuyendo además las limitaciones derivadas del intervalo generacional.

Los principios para la localización de los QTLs son lo mismos que las del análisis de ligamiento entre dos genes mayores. La diferencia es que desconocemos a priori el genotipo para el locus cuantitativo, que consideramos heterocigoto cuando se observen diferencias significativas entre las dos clases fenotípicas, definidas por los alelos alternativos del marcador. Por lo tanto, los elementos necesarios para la detección y localización de QTLs son los mismos que para la elaboración de los mapas de ligamiento entre marcadores, es decir:

- Diseños familiares en los cuales esté segregando el carácter cuantitativo.
- Número suficiente de marcadores que cubran la totalidad del genoma.
- Métodos estadísticos que nos permitan detectar la asociación entre un marcador y un QTL segregante.

Diseños experimentales

Aunque históricamente se han propuesto un gran número de diseños experimentales para la detección de QTLs utilizando marcadores, pueden dividirse en dos grandes grupos, en función de las poblaciones en las que se quiere localizar estos loci (LYNCH y WALSH, 1997):

- Diseños para la localización en "líneas consanguíneas" creadas específicamente para los experimentos o en razas alejadas genéticamente. La localización de QTLs parte del principio de la existencia de dese-

quilibrio de ligamiento entre ellos y el marcador genético. La diferencia fenotípica de las medias entre individuos homocigóticos para el marcador se utiliza como indicadora de la eficiencia del diseño. Para generar desequilibrio de ligamiento es necesario realizar los apareamientos entre líneas parentales que difieren en los genotipos del QTL y de los marcadores, para obtener una F_1 heterocigótica para ambos. A partir de esta generación, se busca la asociación por retrocruzamiento o mediante la producción de una tercera generación segregante F_2.

• Diseños para la localización en poblaciones comerciales o *outbred*. Como tal entendemos poblaciones utilizadas rutinariamente en los sistemas habituales de producción animal. Hacen referencia, fundamentalmente, a los diseños hija y nieta (WELLER *et al.*, 1990); ambos basados en poblaciones de medio hermanos. En el primero, se buscan asociaciones entre las regiones genómicas flanqueadas por marcadores y las producciones de las hijas de varios sementales. Se obtendrán dos grupos de hijas de acuerdo al cromosoma paterno, y a su vez la presencia de alelos alternativos del QTL tenderán a generar diferencias en la media del carácter cuantitativo de cada grupo. WELLER *et al.* (1990) proponen otra alternativa al caso anterior para optimizar el análisis, añadiendo una generación adicional al diseño experimental. Los genotipos se analizan en los hijos y el carácter cuantitativo es medido en las nietas del reproductor y, mediante un test de progenie, se estima el valor genético del hijo para ese carácter.

La utilización de un determinado diseño experimental depende de las particularidades de las poblaciones originarias, del número y tipo de marcadores disponibles y de las características de los fenotipos que se miden en las poblaciones. Normalmente, el diseño elegido es aquel que optimiza la potencia estadística del experimento de detección de QTLs, sin olvidar las limitaciones económicas que suelen conllevar estas experiencias.

La mayoría de los estudios sobre búsqueda de QTLs en animales domésticos, se basan en el cruzamiento de razas alejadas genéticamente, tanto en ganado porcino (cruzamientos entre razas de origen europeo y asiático, o cerdo europeo con jabalí), como en vacuno (cruzamiento entre razas europeas o entre éstas y cebúes) o en avicultura con los cruces entre líneas de ponedoras y de producción de carne (broilers).

En este tipo de cruzamientos, se asume que ambas poblaciones poseen alelos alternativos del QTL, fijados en cada una de las razas (HALEY *et al.*, 1994). La ausencia de variabilidad con respecto al QTL, impide una aplicación directa en los programas de selección que utiliza cada una de las dos líneas implicadas. Su utilización requiere técnicas de introgresión. Estudios recientes, han desarrollado metodologías que no asumen la fijación de los alelos alternativos del QTL, sino frecuencias muy diferentes entre las razas. En consecuencia, varios alelos se encuentran segregando en alguna de las familias, con frecuencias muy distintas, de forma que los QTLs detectados sí podrían ser utilizados directamente en los programas de selección (ALFONSO y HALEY, 1998; PÉREZ-ENCISO y VARONA, 2000).

El diseño de medio hermanos, ya sea de dos (hija) o tres (nieta) generaciones, ha sido aplicado con notable éxito en el ganado vacuno de producción láctea (GEORGES *et al.*, 1995; ZHANG *et al.*, 1998). Los resultados obtenidos están siendo ya utilizados por compañías comerciales en sus programas de selección (BOVENHUIS y SPELMAN, 1998).

Cuadro 1
Tamaño estimado de diferentes experimentos para la localización de QTL

Diseño experimental	Tamaño de la población	Efecto del QTLr
Retrocruzamiento	525-2.100	0,282 s.d
F_2	1.050-2.100 (p = 0,5)	0,282 s.d
Diseño hija	2.000	0,5-1 s.d.
Diseño nieta	1.000	0,2 s.d.

Por lo que se refiere al tamaño teórico de los diseños experimentales, las estimaciones realizadas dependen del tipo de diseño, de la frecuencia de los alelos del locus cuantitativo en la población, de la heredabilidad del carácter y del efecto fenotípico del QTL. En el cuadro 1 se muestran algunas de las estimaciones recogidas de la bibliografía. Como se puede observar el tamaño teórico oscila entre 525 y más de 4.000 animales (SOLLER *et al.*, 1976; WELLER *et al.*, 1990; CARTA y ELSEN, 1999). El tamaño de los experimentos que se están llevando a cabo en la actualidad es bastante variable. En el caso de los diseños de cruzamiento entre razas oscilan entre 500 a 1.000 animales y en los diseños de medio hermanos incluyen alrededor de 1.000 animales.

Metodología estadística

Por lo que respecta a la metodología estadística utilizada en la detección de QTLs, durante la última década se han desarrollado una gran cantidad de procedimientos y algoritmos diferentes. Los primeros fueron diseñados para un solo marcador, sin tener en cuenta la existencia de ligamiento con otros marcadores (SOLLER y GENIZI, 1976; WELLER *et al.*, 1990) y empleaban sistemas de análisis de varianza o máxima verosimilitud. Actualmente, la metodología utilizada se basa en el análisis de intervalos flanqueados por marcadores informativos, que nos permiten conocer el origen del fragmento cromosómico de forma condicionada al genotipo de los dos marcadores flanqueantes. Los métodos estadísticos más frecuentemente empleados son el análisis de regresión (KNOTT *et al.*, 1966; KNOTT *et al.*, 1997) y el de máxima verosimilitud (LANDER y BOTSTEIN, 1989; GEORGES *et al.*, 1995), que producen unos resultados muy similares. También se han desarrollado técnicas que pueden ser utilizadas en variables que no siguen una distribución normal, basadas en el test de rangos de Wilcoxon (COPPIETERS *et al.*, 1998a).

Existen otros tipos de análisis con un gran potencial pero que no están completamente implementados para el análisis rutinario, son los métodos bayesianos y de estimación de componentes de varianza. La estimación de componentes de varianza se ha desarrollado mediante máxima verosimilitud (GOLDAR, 1990) o máxima verosimilitud restringida (GRIGNOLA *et al.*, 1996), que plantean problemas cuando los marcadores no son completamente informativos o cuando existen individuos en el pedigrí sin dato genotípico.

Los métodos bayesianos presentan una serie de ventajas claras, como la posibilidad de incluir el número de QTLs en cada cromosoma como parámetro no conocido, la de

tener en cuenta todas las relaciones existentes entre los animales o la de obtener parámetros menos sesgados en las poblaciones sometidas a selección (HOESCHELE *et al.*, 1997). Se ha aplicado con algún problema a pedigríes complejos (HOESCHELE *et al.*, 1997), aunque los avances en esta metodología podrían hacer que se convierta en una de las rutinarias en los próximos años (BINK y VAN ARENDONK, 2000; GEORGE *et al.*, 2000).

Un aspecto a tener en cuenta, es el cálculo de los niveles de significación. La realización de un elevado número de test estadísticos a lo largo del genoma, para buscar asociaciones positivas entre los marcadores y los caracteres objeto de estudio, hace que la cantidad de falsos positivos que se pueden encontrar sea muy elevada, aunque "α" sea muy pequeña en cada uno de los tests. Los procedimientos más frecuentes para fijar los niveles de significación son los del cálculo empírico para cada carácter mediante permutaciones (CHURCHILL y DOERGE, 1994) o la utilización de la corrección de Bonferroni para tests múltiples (Lynch y Walsh, 1997).

Marcadores genéticos

Además de los diseños familiares y los métodos estadísticos, el último aspecto fundamental en la detección de QTLs son los marcadores genéticos. Los primeros estudios realizados emplearon marcadores morfológicos y posteriormente grupos sanguíneos y polimorfismos bioquímicos (NEIMANN-SORENSEN y ROBERTSON, 1961). Pero no fue hasta la llamada "revolución genética", que aplicó la tecnología del DNA recombinante, cuando se dispuso de un número suficiente de marcadores que permitieran una búsqueda sistemática en el genoma de los genes que controlan un carácter cuantitativo, empleando el denominado *genome scan.*

Los primeros mapas de ligamiento realizados en mamíferos empleaban los RFLPs como marcadores, pero fue a partir del descubrimiento de la PCR y los microsatélites, cuando es posible obtener mapas genéticos densos de la práctica totalidad de especies domésticas (ARCHIBALD *et al.*, 1995; KAPPES *et al.*, 1997; CRAWFORD *et al.*, 1996) y se empiezan a desarrollar los primeros proyectos de detección y localización de QTLs en estos animales. En la actualidad, los polimorfismos del tipo SNPs (snips), producidos por una mutación puntual en la secuencia de DNA, parecen ser los marcadores que sustituirán en un futuro cercano a los microsatélites. Este último tipo de marcadores está siendo utilizado con gran éxito, tanto en el hombre como en el ratón, merced a su elevado número y distribución uniforme a lo largo del genoma, y a la facilidad de automatización para su identificación, basada en la tecnología de micropunteo (*microarray*) o fotoactivación que permite aplicar los denominados DNA-chips (LINDBLAD-TOH *et al.*, 2000). En los animales domésticos se encuentran en sus primeras etapas de desarrollo y empiezan a surgir los primeros prototipos de marcadores que cubran el conjunto del genoma en diversas especies, para la posterior producción de los DNA-chips.

En el ámbito de la producción animal, casi todos los estudios realizados o en proceso para la detección de caracteres complejos emplean los microsatélites como marcadores. Los microsatélites son secuencias simples de DNA repetido. Son simples debido a que el tamaño de la unidad iterativa es pequeña (de 2 a 5 nucleótidos). Son muy abundantes en el genoma y se encuentran

distribuidos de una manera prácticamente uniforme a lo largo del mismo, poseen un elevado grado de polimorfismo, presentan herencia codominante y su detección puede automatizarse. Son estas características las que los convierten en marcadores casi ideales. Para poder aplicar el *genome scan* suele utilizarse un marcador polimórfico cada 20 cM. Con objeto de cubrir la totalidad del genoma se emplean, en las especies de animales domésticos, una media de 180 a 200 marcadores en una primera aproximación. Si tenemos en cuenta el tamaño de los diseños experimentales comentado con anterioridad, la realización de un *genome scan* implica la producción de más de 100.000 datos genotípicos.

Esta enorme demanda de trabajo rutinario hace necesario aplicar técnicas de alto rendimiento por persona y año. En un principio se utilizaron técnicas de amplificación conjunta (*multiplex*) utilizando marcado radiactivo y lectura de los alelos como valores relativos. Posteriormente, se han utilizado diversas tecnologías derivadas de la secuenciación automática, aunque parece que las técnicas fluorescentes basadas en la detección de 4 colores por línea de corrido electroforético desarrolladas por ABI y aplicadas posteriormente por otras firmas, son las más utilizadas en los proyectos de localización de QTLs que emplean microsatélites como marcadores (ZIEGLE *et al.*, 1992; SCHWENGEL *et al.*, 1994; HALL *et al.*, 1996).

El volumen de trabajo laboratorial necesario y la fiabilidad de la identificación genotípica de cada marcador que requiere el diseño experimental, hace que deba prestarse una atención especial a la metodología empleada. Brevemente, pasaremos a describir el proceso de identificación genotípica con la metodología de cuatro colores por línea utilizando secuenciadores automáticos. Los microsatélites se amplifican mediante PCR-multiplex de forma que con cada fluorocromo se amplifican de 2 a 4 marcadores. Una vez amplificados, se mezclan de forma proporcional, para que no exista contaminación cromática entre ellos y para que la intensidad sea similar. Finalmente, uno de los fluorocromos se reserva para el estándar de tamaño que servirá para la detección de la longitud de los alelos de cada animal. La mezcla de marcadores se separa mediante electroforesis en condiciones desnaturalizantes, en un secuenciador automático. Los datos de intensidad lumínica y de tiempo se registran en un ordenador. Finalizada la electroforesis, los datos brutos se analizan mediante un software diseñado específicamente para ello (GeneScan) y así la señal policromática detectada por el secuenciador se divide en cuatro señales monocromáticas. A partir del estándar se calcula el tamaño de los picos de los marcadores en los ficheros de resultados. A continuación, cada uno de los picos detectados se agrupa dentro de un marcador determinado y se le asigna una "etiqueta alélica", que es la que se utilizará posteriormente en los diferentes tipos de análisis.

Aunque el proceso es aparentemente sencillo y no parece presentar demasiadas fuentes de error, hay que tener en cuenta que, unido a esta potente tecnología encontramos una serie de problemas, a menudo no considerados a la hora de comenzar un proyecto de identificación de genotipos a gran escala, que pueden dar lugar a errores suficientes como para dificultar la correcta detección, localización y estimación de la magnitud del efecto de genes con influencia sobre un carácter cuantitativo (BRZUSTOWICZ *et al.*, 1993).

Entre los tipos de errores más frecuentemente detectados en nuestro laboratorio, podemos destacar los siguientes:

• Bandas sombra que dificultan la identificación genotípica de forma automática. Son específicas de cada secuencia amplificada y se pueden atenuar mediante cambios en las condiciones de PCR, aunque en algunas secuencias son difíciles de evitar (KNOWLES *et al.*, 1992).

• En algunos casos se observa una amplificación diferencial entre los dos alelos de un locus. Normalmente se debe a la existencia de una mutación en la zona complementaria de uno de los dos "primers" y puede determinar un primer grado en la aparición de alelos nulos, en los cuales uno de los dos alelos de los microsatélites no amplifica. En este caso, el marcador se trata como si tuviese un modo de herencia dominante (CALLEN *et al.*, 1993).

• La actividad "+A" de la Taq polimerasa se ha observado con cierta frecuencia y en algunas ocasiones ocasiona una diferente longitud del fragmento amplificado, lo que determina una mayor varianza en las clases alélicas e, incluso, un solapamiento entre las mismas (SMITH *et al.*, 1995).

• Fallos en la asignación del estándar interno. Esto conlleva que se realice una estimación diferente en el tamaño de los alelos, entre las diferentes líneas de un gel o entre geles distintos.

• Anomalías electroforéticas. Suele ser debidas a una migración electroforética errática de algún tipo cadenas de DNA, principalmente aquellas de naturaleza repetida. Ya que la estimación de la longitud alélica se realiza por comparación con un fragmento de tamaño de DNA conocido y de naturaleza no repetida y como resultado a un mismo alelo se le asignan longitudes distintas en geles diferentes (GHOSH *et al.*, 1997).

Por lo que se refiere a este último tipo, en un estudio realizado con medidas repetidas de un mismo individuo en geles distintos, se han identificado los siguientes factores de variación en la asignación alélica: gel, línea dentro del gel y, fundamentalmente, la interacción entre gel y marcador. En el cuadro 2 se resumen los factores de variación analizados y el porcentaje de varianza que explica cada uno de ellos. Como se puede observar, la combinación "Gel x Marcador" explica el mayor porcentaje de la varianza detectada en la asignación alélica. Estos resultados parecen indicar que ciertas secuencias microsatélites presentan una menor repetibilidad en el cálculo del tamaño de sus alelos, sin duda debido a que su composición en pares de bases hace que la migración electroforética sea más variable, al ser comparada con un estándar de tamaño compuesto por DNA de naturaleza no repetida.

Cuadro 2
Factores de variación en la asignación del tamaño de los alelos

	Significación	Varianza explicada
Gel	***	21%
Línea (gel)	***	3%
Gel x Marcador	***	68%
Alelo (marcador)	NS	

Estrategias alternativas

Existen una serie de estrategias que permiten reducir, de una manera significativa, el número de datos genotípicos necesarios para abordar una búsqueda de QTLs, sin que exista una pérdida notable de precisión. Estos procedimientos son interesantes desde el punto de vista de la reducción de costes laboratoriales y del tiempo necesario para la finalización del proyecto. El aspecto de la reducción del presupuesto es de particular relevancia, ya que la producción de un número tan elevado de datos genotípicos como los que se requieren para este tipo de proyectos, lleva asociado un alto coste en reactivos y personal, normalmente el más importante del proyecto, ya que la medición de los caracteres fenotípicos suele resultar menos gravosa. Entre dichas estrategias podemos destacar el "muestreo selectivo" y el *DNA pooling*.

El denominado muestreo selectivo o *selective genotyping* (Lander y Botstein, 1989) consiste en la elección para su análisis laboratorial, de los individuos con fenotipos extremos para el carácter que se desea analizar. Los fenotipos se adaptan mayoritariamente a una distribución normal y se suelen elegir entre el 40 y 50% de los individuos de la población, tomados de ambas colas de la distribución fenotípica. DARVASI y SOLLER (1992) demuestran que el genotipado selectivo puede determinar una considerable reducción del número de individuos analizados, sin disminuir de una forma considerable la potencia estadística del experimento. Este tipo de estrategia presenta como inconveniente principal que no es muy útil en casos de análisis para más de un carácter, ya que individuos con fenotipos extremos para uno pueden no serlo para otro.

El número de animales analizados se puede reducir también por medio de la mezcla de DNA (DNA pooling o Bulked segregant analysis). Es decir, en lugar de identificar los genotipos individualmente, el material genético de los individuos con similares fenotipos (alto y bajo), son separados en dos mezclas. De este modo, cada marcador podrá ser examinado mediante dos únicos análisis, uno por cada mezcla, de acuerdo a la intensidad de la banda que producen (DARVASI y SOLLER, 1994). Obviamente, los genotipos individuales para cada marcador no pueden ser determinados, pero sí pueden predecirse las frecuencias alélicas relativas, mediante un proceso de análisis densitométrico de las mezclas de ADN amplificado (KHATIB *et al.*, 1994).

Debido al error técnico que se comete en la determinación de las frecuencias alélicas, por la presencia de bandas superpuestas, es menos potente que el método anterior, pero resulta útil combinarlo con el método de muestreo selectivo, para el estudio en conjunto de diferentes caracteres. Un análisis de este tipo se ha aplicado, recientemente, a la búsqueda de QTLs en ganado vacuno lechero (LIPKIN *et al.*, 1998).

Otra estrategia de optimización del análisis de QTLs, aunque desde el punto de vista fenotípico, es la aplicación del "test de progenie". GEORGES *et al.* (1995) utilizan un diseño abuelo, donde el análisis de los genotipos se realiza en los hijos resultantes del apareamiento entre padres con méritos genéticos superiores ya probados, y el valor del carácter cuantitativo se obtiene mediante un "test de progenie", en función del comportamiento de sus hijas para el carácter de interés.

Análisis de los resultados

Una vez finalizada la determinación de genotipos se realiza la construcción del mapa de ligamiento y finalmente el análisis estadístico.

Los resultados que se obtienen en un *genome scan* suelen dar un intervalo de confianza de localización de los QTLs de unos 30-40 cM, demasiado grande para el inicio de estrategias de clonado posicional, e incluso poco eficiente para la utilización en MAS.

El primer paso para acortar el intervalo de confianza para la localización de un QTL es el aumento de densidad de marcadores. Dado que actualmente la mayoría de los mapas de ligamiento de las especies domésticas poseen una densidad media o elevada, se deberían utilizar los marcadores existentes en el intervalo de confianza de la localización del QTL. Si no fuesen suficientes, habría que recurrir a la producción de nuevos marcadores. La estrategia más sencilla sería la localización de nuevos microsatélites o SNPs en YAC o BAC que contienen a los marcadores del intervalo de confianza.

Por lo que se refiere al aumento de la densidad de recombinantes, en animales domésticos las estrategias diseñadas para animales de laboratorio (producción de líneas recombinantes, RIL, AIL, RPT, ISCS, etc.) (DARVASI, 1998) son poco eficaces debido al coste y al intervalo generacional. Por lo tanto, la única posibilidad de incrementar el número de recombinantes sería el aumento de tamaño del pedigrí a analizar. Otra posibilidad sería la utilización del desequilibrio de ligamiento existente en algunas poblaciones como es el caso del ganado vacuno de raza frisona (FARNIR *et al.*, 2000). Este método se ha mostrado muy útil en los caracteres monogénicos (CHARLIER *et al.*, 1996; HOUWEN *et al.*, 1994; DUNNER *et al.*, 1997) y con la misma base teórica puede ser aplicado a los caracteres cuantitativos. El principio se basa en que los individuos portadores de un mismo alelo "raro" descienden de un único individuo en el que se produjo la mutación, y portan un fragmento cromosómico común, cuya longitud es igual a 2/(n*x), siendo n el número de generaciones que separan el individuo fundador y el actual y "x" la longitud en cM.

Aplicación al cromosoma 14 del ganado vacuno

En un estudio de búsqueda de QTLs mediante *genome scan*, en ganado vacuno, utilizando un diseño abuelo, se detectó un fuerte efecto para el porcentaje de grasa, en la región proximal del cromosoma 14, que fue confirmado en una muestra independiente de animales (COPPIETERS *et al.*, 1998b). El intervalo de confianza para la localización de dicho QTL era de aproximadamente 40 cM, evidentemente demasiado grande para la realización de una aplicación directa en MAS o para un posible clonado posicional (10 cM $\approx$ 10 x 10^6 bp $\approx$ 300 genes). El proceso de "mapeo fino" de dicho QTL se presenta en RIQUET *et al.* (1999). Con el fin de lograr una localización más precisa de dicho QTL se siguieron los pasos siguientes:

1. Obtención de un mapa de alta resolución en la región proximal del cromosoma 14. La densificación en marcadores de la zona se realizó mediante la búsqueda de microsatélites y SNPs en BAC y YAC que contenían los marcadores utilizados en el primer análisis.

2. Identificación de las familas segregantes para el QTL. Las familias que superaron el nivel de significación del 5% (corregido mediante el procedimiento de Bonferroni) fueron seleccionadas como portadoras del QTL y del correspondiente fragmento circundante. En este caso se seleccionaron siete de los 29 sementales analizados como heterocigotos para dicho QTL.

3. Determinación de los genotipos para los nuevos marcadores producidos y determinación de la fase de ligamiento de los mismos en los sementales portadores del QTL.

4. Identificación de un haplotipo compartido en uno de los dos alelos de los sementales. Este haplotipo compartido por el conjunto de hijos portadores del alelo "+" o del "-" será el portador del QTL, y el tamaño del mismo dependerá del número de generaciones transcurridas entre los sementales tipificados y el antepasado común.

5. Finalmente, se verificó el efecto fenotípico del segmento portador del alelo del QTL en la población general. Para ello se realizó un análisis de las medias productivas de los individuos portadores de dicho haplotipo frente al resto de haplotipos. El resultado demuestra claramente que los individuos portadores del haplotipo "+" poseen una media fenotípica superior a los no portadores.

Mediante este procedimiento se caracterizó un segmento cromosómico de 5 cM, donde se ha localizado un QTL con efecto positivo sobre el porcentaje de grasa en la leche. El tamaño de este fragmento permite que pueda ser utilizado en selección asistida por marcadores y es el primer paso para la realización del posible clonado posicional.

El paso siguiente es el clonado posicional del QTL detectado, para ello se pueden utilizar diversas técnicas como el secuenciado de la región completa, *el zoo-blotting, exon trapping* o la utilización del mapeo comparado que permite la localización de genes candidatos posicionales. De todas formas, el trabajo requerido para el clonado posicional es muy laborioso y dilatado en el tiempo. Una vez identificado el gen deben caracterizarse su productos y regulación mediante técnicas de análisis de su expresión y de las proteínas producidas por el mismo.

Bibliografía

ALFONSO L., HALEY C.S., 1998. Power of different F2 schemes for QTL detection in livestock. Anim Sci; 66: 1-8.

ARCHIBALD A.L., HALEY C.S., BROWN J.F., COUPERWHITE S., MCQUEEN H.A., NICHOLSON D., COPPIETERS W., VAN DE WEGHE A., STRATIL A., WINTERO A.K., *et al.*, 1995. The PiGMaP consortium linkage map of the pig (Sus scrofa). Mamm Genome; 6: 157-175.

BRZUSTOWICZ L.M., MERETTE C., XIE X., TOWNSEND L., GILLIAM T.C., OTT J., 1993. Molecular and statistical approaches to the detection and correction of errors in genotype databases. Am J Hum Genet; 53: 1137-1145.

CARTA, A., ELSEN, J.M., 1999. Sire design power calculation for QTL mapping experiments. Genet Sel Evol, 31: 183-189.

CHARLIER C., FARNIR F., BERZI P., VANMANSHOVEN P., BROUWERS B., GEORGES M., 1996. IBD mapping of recessive traits in livestock: application to map the bovine syndactyly locus to chromosome 15. Genome Res; 6: 580-589.

CHURCHILL G.A., DOERGE R.W., 1995. Empirical threshold values for quantitative trait mapping. Genetics, 138: 963-971.

COPPIETERS W., KVASZ A., FARNIR F., ARRANZ J.J., GRISART B., MACKINNON M., GEORGES M., 1998b. A Rank-based nonparametric method for mapping trait loci in outbred half-sib pedigrees: application to milk production in a grand-daughter design. Genetics, 149: 1547-1555.

COPPIETERS W., RIQUET J., ARRANZ J.J., BERZI P., CAMBISANO N., GRISART B., KARIM L., MARCQ F., MOREAU L., NEZER C., SIMON P., VANMANSHOVEN P., WAGENNAR D., GEORGES M, 1998a. A QTL with major effect on milk yield and composition maps to bovine chromosome 14. Mamm Genome, 9: 540-544.

CRAWFORD A.M., DODDS K.G., EDE A.J., PIERSON C.A., MONTGOMERY G.W., GARMONSWAY H.G., BEATTIE A.E., DAVIES K., MADDOX J.F., KAPPES S.W. *et al.*, 1995. An autosomal genetic linkage map of the sheep genome. Genetics; 140: 703-724.

DARVASI A., SOLLER M., 1992. Selective genotyping for determination of linkage between a marker locus

and a quantitative trait locus. Theor Appl Genet; 85: 353-359.

DARVASI A., SOLLER M., 1994. Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics; 138: 1365-1373.

DARVASI A., 1998. Experimental strategies for the genetic dissection of complex traits in animal models. Nat Genet; 18:19-24.

DUNNER S., CHARLIER C., FARNIR F., BROUWERS B., CANON J., GEORGES M., 1997. Towards interbreed IBD fine mapping of the mh locus: double-muscling in the Asturiana de los Valles breed involves the same locus as in the Belgian Blue cattle breed. Mamm Genome; 8: 430-435.

GEORGE A.W., MENGERSEN K.L., DAVIS G.P., 2000. Localization of a quantitative trait locus via a Bayesian approach. Biometrics; 56: 40-51.

GEORGES M., NIELSEN D., MACKINNON M., MISHRA A., OKIMOTO R., PASQUINO A.T., SARGEANT L.S., SORENSEN A., STEELE M.R., ZHAO X., WOMACK J.E., HOESCHELE I., 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics. 139: 907-920.

GHOSH S., KARANJAWALA Z.E., HAUSER E.R., ALLY D., KNAPP J.I., RAYMAN J.B., MUSICK A., TANNENBAUM J., TE C., SHAPIRO S., ELDRIDGE W., MUSICK T., MARTIN C. *et al.*, 1997. Methods for precise sizing, automated binning of alleles, and reduction of error rates in large-scale genotyping using fluorescently labeled dinucleotide markers. FUSION (Finland-U.S. Investigation of NIDDM Genetics) Study Group. Genome Res.; 7: 165-178.

GOLDGAR D.E., 1990. Multipoint analysis of human quantitative genetic variation. Am J Hum Genet; 47: 957-967.

GRIGNOLA F.E., HOESCHELE I. Y TIER B., 1996. Mapping quantitative trait loci via residual maximum likelihood: I Methodology. Genet Sel Evol; 28: 479-490.

HALL J.M., LEDUC C.A., WATSON A.R., ROTER A.H., 1996. An approach to high-throughput genotyping. Genome Res. 6: 781-790.

HOESCHELE I., UIMARI P., GRIGNOLA F.E., ZHANG Q., GAGE K.M., 1997. Advances in statistical methods to map quantitative trait loci in outbred populations. Genetics; 147: 1445-1557.

HOUWEN R.H., BAHARLOO S., BLANKENSHIP K., RAEYMAEKERS P., JUYN J., SANDKUIJL L.A., FREIMER N.B., 1995. Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. Nat Genet.; 8: 380-986.

KAPPES S.M., KEELE J.W., STONE R.T., MCGRAW R.A., SONSTEGARD T.S., SMITH T.P.L., LOPEZ-CORRALES N.L., BEATTIE C.W., 1997. A second-generation linkage map of the bovine genome. Genome Res. 7: 235-249.

KHATIB H., DARVASI A., PLOTSKI Y., SOLLER M., 1994. Determining relative microsatellite allele frequencies in pooled DNA samples. PCR Methods and Applications. 4: 13-18.

KNOTT S.A., ELSEN J.M. Y HALEY C.S., 1996. Methods for multiple marker mapping of quantitative trait loci in half-sib populations. Theor Appl Genet; 93: 71-80.

KNOTT S.A., NEALE D.B., SEWELL M.M. Y HALEY C.S., 1997. Multiple marker mapping of quantitative trait loci in an outbred pedigree of loblolly pine. Theor Appl Genet; 94: 810-820.

KNOWLES J.A., VIELAND V.J., GILLIAM T.C., 1992. Perils of gene mapping with microsatellite markers. Am J Hum Genet.; 51: 905-909.

LANDER E.S., BOTSTEIN, D., 1989. Mapping mendelian factor underlying quantitative traits using RFLP linkage map. Genetics, 121: 185-199.

LINDBLAD-TOH K., WINCHESTER E., DALY M.J., WANG D.G., HIRSCHHORN J.N., LAVIOLETTE J.P., ARDLIE K., REICH D.E., ROBINSON E., SKLAR P., SHAH N., THOMAS D., FAN J.B., GINGERAS T., WARRINGTON J., PATIL N., HUDSON T.J., LANDER E.S., 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nat Genet; 24:381-386.

LIPKIN E., MOSIG M.O., DARVASI A., EZRA E., SHALOM A., FRIEDMANN A., SOLLER M., 1998. Quantitative Trait Locus mapping in dairy cattle by means of selective milk dna pooling using dinucleotide microsatellite markers: analysis of milk protein percentage Genetics; 149: 1557-1567.

LYNCH M., WALSH B., 1997. Genetics and Analysis of Quantitative Traits. Sinauer Associates Inc. Publishers. Sunderlan, Massachusset, USA.

NEJMANN-SORENSEN A., ROBERTSON A., 1961. The association between blood groups and several produc-

tion characteristics in three Danish cattle breeds. Acta Agric Scan, 11: 163-196.

PÉREZ-ENCISO M., VARONA L., 2000. Quantitative Trait Loci Mapping in F2 Crosses Between Outbred Lines. Genetics; 155: 391-405.

RIQUET J., COPPIETERS W., CAMBISANO N., ARRANZ J.J., BERZI P., DAVIS S.K., GRISART B., FARNIR F., KARIM L., MNI M., SIMON P., TAYLOR J.F., VANMANSHOVEN P., WAGENAAR D., WOMACK J.E., GEORGES M., 1999. Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle. Proc Natl Acad Sci U S A; 96: 9252-9257.

SAX K. (1923). The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris. Genetics. 8: 552-560.

SCHWENGEL D.A., JEDLICKA A.E., NANTHAKUMAR E.J., WEBER J.L., LEVITT R.C., 1994. Comparison of fluorescence-based semi-automated genotyping of multiple microsatellite loci with autoradiographic techniques. Genomics; 22: 46-54.

SMITH J.R., CARPTEN J.D., BROWNSTEIN M.J., GHOSH S., MAGNUSON V.L., GILBERT D.A., TRENT J.M., COLLINS F.S., 1995. Approach to genotyping errors caused by nontemplated nucleotide addition by Taq DNA polymerase. Genome Res.; 5: 312-317.

SOLLER M., GENIZI A., 1978. The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. Biometrics. 34: 47-55.

SOLLER M., GENIZI A., BRODY T., 1976. On power of experimental designs for detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor Appl Genet. 47: 35-39.

SPELMAN R.J. Y BOVENHUIS H., 1998. Moving from QTL experimental results to the utilization of QTL in breeding programmes. Anim Genet. 29: 77-84.

WELLER J.I., KASHI Y., SOLLER M., 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J Dairy Sci.; 73: 2525-2537.

ZHANG Q., BOICHARD D., HOESCHELE I., ERNST C., EGGEN A., MURKVE B., PFISTER-GENSKOW M., WITTE L.A., GRIGNOLA F.E., UIMARI P., THALLER G., BISHOP M.D., 1998. Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. Genetics; 149: 1959-1973.

ZIEGLE J.S., SU Y., CORCORAN, K.P., 1992. Application of automatic DNA sizing for genotyping microsatellite loci. Genomics, 4: 1026-1031

ITEA (2000), Vol. **96A N.° 3,** 184-188

RePed. A TOOL FOR CHECKING, EXPLORING AND DEBUGGING PEDIGREES

J.A. Baro de la Fuente, R. Álvarez, C.E. Carleos, D. García, H. Lamelas

Introduction

The purpose of this paper is to present a software tool that analyses large pedigrees formatted as data files with an individual-father-mother structure.

The main problems addressed were the complexity of family links and memory allocation. This large pedigree files also need an efficient and informative error test. Finally, some other utilities were added to deal with subpedigrees and genotypes.

Methods

RePed is a tool developed for checking, exploring and debugging pedigrees. RePed tasks are performed through use of a structured type consisting on an integer for the numerical identity of the animal (automatically set by the program), a string of characters for its alphanumerical identity, a real value for the inbreeding coefficient and two pointers to strutured type for the sire and dam of the animal. This pointer structure provides an interpretation of the pedigree as an oriented graph, defined as the graph formed by the union of a set of oriented binary trees.\\

There are two types of input files for RePed:

1. Pedigree files: These are files with at least three alphanumerical columns: the alphanumerical identity of an animal, of its sire and of its dam. A fourth column can be added containing the animal genotype.

2. Data files: These are files with the alphanumerical identity of an animal in the first column and the second column may contain the alphanumerical identity of the individual's dam.

After reading all data, non-informative animals are deleted from pedigree. An animal is considered non-informative when it is repeated or when it is a generic, unknown animal.

Animals are allocated within the structured vector by first piling up the three identity columns and simultaneously fetching each individual's sire and dam.

Error detection and correction

When all animals are stored in the structured vector, a pedigree depuration is performed to detect the following types of errors:

1. Non consistent records: Repeated animals with different sets of parents, self-sire, or self-dam individuals.

2. Sex errors: An animal is found in both sire's column and dam one.

3. Circular pedigree: A circular pedigree error is found when an animal appears as its own ancestor.

Non consistent records and sex errors may be automatically corrected and the pedigree file may be overwritten with a corrected one. There is a choice of implemented correction policies: suspicious animals are prefixed by 0, or substituted by generic animals.

Circular pedigree errors cannot be automatically corrected. As this error must be corrected to avoid infinite loops during the execution of the program, an ascii-art subtree of the circular path is written on an error file, to guide the user for manual correction.

Each error condition (corrected or not) is documented and collected in a detailed report file.

When a pedigree has no errors, it may be analysed using genetic diversity coefficients. Implemented coefficients are imbreeding coefficient F, effective number of founders f_e, effective number of ancestors f_a and average relatedness AR.

Inbreeding coefficient

The inbreeding coefficient may be defined as the probability of being identical-by-descent (IBD) homozygous. This may be achieved by use of the subpedigree containing all of an individual's ancestors, i.e., the oriented subtree that has this individual as vertex.

There are two methods to calculate the inbreeding coefficient:

1. First one involves searching for all possible paths between an animal's sire i_s and dam i_d passing through a common ancestor of i_s and i_d AN. Only paths verifying that the intersection of the set formed by animals in the path from i_s to AN path and the set formed by animals in the path from i_d to AN path is the common ancestor AN will be considered as valid ones.

Using graph theory notation:

$$AN \in \left(\Gamma(i_s) \cup \Gamma^2(i_s) \cup ... \cup \Gamma^{n_s}(i_s)\right) \cap \left(\Gamma(i_d) \cup \Gamma^2(i_d) \cup ... \cup \Gamma^{n_s}(i_d)\right),$$

with $\Gamma(i) = \{i_s, i_d\} = \{\text{parents of } i \text{ vertex}\} = \{\text{vertex following } i\}$.

The contribution to the inbreeding coefficient of each path takes the value $\left(\frac{1}{2}\right)^{n_j}\left(1 + F(AN(j))\right)$, where $F(AN(j))$ is the inbreeding coefficient of the j-th common ancestor of i_s and i_d and n_j is the total number of vertices in i_s to i_d path through $AN(j)$, i.e., $n_j = \#C_{i_s,AN(j)} + \#C_{i_d,AN(j)}$, with $C_{i,j}$ being the path between vertex i and vertex j.

Inbreeding coefficient is the sum over all common ascendents of this partial contributions.

2. The second method is based on coancestry. Letting A and B be i_s parents and C and D be i_d ones (A, B, C and D are the 4 grandparents of selected animal i), then the inbreeding coefficient of animal i will be

$$F(i) = f(i_s, i_d) = \tfrac{1}{4} f(A,C) + \tfrac{1}{4} f(A,D) + \tfrac{1}{4} f(B,C) + \tfrac{1}{4} f(B,D).$$

This second method allows an easy implementation, but requires the construction of the coancestry matrix and memory allocation for large pedigree files.

RePed follows the first method. The strategy is as follows:

1. Firstly all common ancestors of each couple (i_s, i_d) have to be found. Let $\$V_{i,k}$ be the intersection of i's ancestors and k's descendents, i.e.

$$V_{i,k} = \left\{ j / j \in \left(\Gamma^1(i) \cup \Gamma^2(i) \cup ... \cup \Gamma^{n_i}(i)\right) \cap \left(\Gamma^{-1}(k) \cup \Gamma^{-2}(k) \cup ... \cup \Gamma^{n_k}(k)\right) \right\}.$$

2. Let $V_{AN(j)} = V_{is.AN(j)} \cup V_{id.AN(j)}$ be the binary subtree of interesting vertices related to ancestor $AN(j)$. Then set

$$V = \bigcup_{AN(j)} V_{AN(j)}$$

will be the set of vertices for obtaining the inbreeding coefficient of individual i. All superflous and non informative branches have been deleted.

3. Now, $AN(j)$'s inbreeding coefficient must be evaluated $\forall j$. If they are not calculated yet, make a recursive call to inbreeding coefficient calculation's subroutine.

4. Not all paths between i_s to i_d through common ancestor $AN(j)$ are pertinent. Only paths verifying $C_{i_s.AN(j)} \cap C_{i_d.AN(j)} = AN(j)\}$ are required.

Finally, the inbreeding coefficient will be obtained as:

$$F(i) = \sum_{AN(j)} \sum_{C \wp(V_{AN(j)}) C_{i_s.AN(j)}} \sum_{C \wp(V_{AN(j)})} \left(\frac{1}{2}\right)^{\#C_{i_s.AN(j)} + \#C_{i_d.AN(j)}} \left(1 + F(AN(j))\right)$$

Effective number of founders

Effective number of founders is a genetic diversity coefficient (MEUWISSEN and LUO 1992) defined as the number of equally contributing founders that would be expected to produce the same genetic diversity as in the population under study. A founder is defined as every individual with unknown parents in the population under study.

Any gene randomly sampled at any autosomal locus of a given individual has a probability $\frac{1}{2}$ of originating from its sire and a probability $\frac{1}{2}$ of originating from its dam, and, in general, a probability $(\frac{1}{2})^n$ of originating from one animal in $\Gamma^n(i)$ (animals on the n-th previous generation). Under this hypothesis a probability of gene origin $q(i)$ can be calculated for every animal i in the pedigree.

The sum over all population founders of the probability of gene fulfills

$$\sum_{\{i / i \text{ founder}\}} q(i) = 1.$$

The coefficient proposed by LACY and ROCHAMBEAU for genetic variability measurement was

$$f_e = \frac{1}{\sum_{i \in B} q(i)^2} \quad \text{with } B = \{k / \Gamma(k) = \varnothing\}$$

The algorithm used for calculating the probabilities of gene origin $q(i)$ (BOICHARD *et al.*) are detailed below:

1. Define the population under study, i.e. the group of N animals carrying the gene pool of interest.

2. Initialize a vector q with 1 for animals in the population under study and with 0 otherwise.

3. Process pedigree file counter generation wise

$$q(k) = q(k) + \tfrac{1}{2} q(i), \ \forall k \in \Gamma(i)$$

4. Divide vector q by the population size N

Finally, $f_e = \frac{1}{\sum_{i \in B} q(i)^2}$ with $B = \{k / \Gamma(k) = \varnothing\}$

Effective number of ancestors

An important limitation of the effective number of founders is that it ignores potential bottlenecks in the pedigree, i.e., it does not take account of the possibility of a popu-

lation formed by fullsibs of unrelated parents. In such case, f_e will be equal to 2^n if n-th previous generation were considered although genetic diversity may be explained by only two individuals.

The effective number of ancestors coefficient tries to overcome this problem looking for the minimum number of ancestors (founders or not) needed to explain the genetic diversity observed in the population.

Ancestors are selected on the basis of their expected genetic contribution. As ancestors may not be founders, the may be related, so the sum of probabilities of gene origin $\boldsymbol{q}$ over all founders may be greater than 1, and indication that some redundant information must be eliminated.

A new probability vector must be defined, $\boldsymbol{p}$, such that $p(i)$ is the contribution of individual i to genetic diversity not yet explained by the other ancestors (marginal contribution of the ancestor i).

Probability vector $\boldsymbol{p}$ could be derived from $\boldsymbol{q}$, eliminating redundancies.

The algorithm used for the calculation of the effective number of ancestors is the following:

1. Define the population under study, i.e., the group of N animals carrying the gene pool of interest.

2. Assume that first k-1 most important ancestors are already found.

3. Delete pedigree information (sire and dam information) for the k-1 ancestors already found.

4. Initialize a vector q with 1 for animals in the population under study and with 0 otherwise

5. Initialize another vector a with 1 for the k-1 already selected ancestors and with 0 otherwise.

6. Process the pedigree counter-generation-wise

$$q(k) = q(k) + \tfrac{1}{2}\, q(i),\ \forall k \in \Gamma(i)$$

7. Process the file generation-wise

$$a(i) = a(i) + \tfrac{1}{2}\, a(k),\ \forall k \in \Gamma(i)$$

8. Compute the marginal contribution $p(i)$ by substracting the contributions of the ancestors already selected: $p(i) = q(i)\big(1\text{-a(i)}\big)$.

9. Select the k-th ancestor: $\$i_k \,/\, p(i_k) > p(i),\ \forall i$

10. Go to 3.

Finally divide vector p by N. Then the effective number of ancestors will be

$$f_a = \frac{1}{\sum_{i \in B} p(i)^2} \text{ with } B=\{k/\ \Gamma(k)=\varnothing\}$$

Average relatedness

Average relatedness (AR) is a genetic diversity coefficient as defined by J.P. GUTIÉRREZ and J. CAÑÓN (personal communication). Average realtedness takes account not only of coancestry but also of inbreeding, so it can be used either as a genetic variability index or as a measurement to compare inbreeding between two populations.

A vector c containing average relatedness coefficients is defined as

$$c' = \tfrac{1}{n}\ 1'A,$$

with A being the coancestry matrix.

The algorithm to obtain c is shown below:

1. Initialize c = 1

2. Process pedigree file counter generation wise

$$c(k) = c(k) + \tfrac{1}{2}\ c(i),\ \forall k \in \Gamma(i)$$

3. Divide c by the population size

If i is a founder, $AR(i) = c(i)$. But if i is an ancestor but not a founder, we need to relate the i-th individual with the rest of the population except its progeny. This can be done by adding $c(i)$ half of the "free" contribution of its parents:

$$c(i) = c(i) + \tfrac{1}{2}\left[c(i_s) - \tfrac{1}{2}\left(1 + F(i_s)\right) c(i)\right] + \\ + \tfrac{1}{2}\left[c(i_d) - \tfrac{1}{2}\left(1 + F(i_d)\right) c(i)\right],$$

where $F(j)$ is the inbreeding coefficient of the j-th animal.

Other utilities

RePed includes some other utilities.

The program may create subpedigrees for a given data set (read from a data file) and a master pedigree file by recursively including all ancestors of the animals on the data set to an *animal-sire-dam* structure. A pedigree created this way is the minimal complete pedigree containing all data file individuals.

RePed can also recodify pedigree and data files, subtituting alphanumeric identities by numeric ones. This numeric identities are automaticaly set in a generation wise-order. This assingment method assures that every ancestor of an animal will have a numerical identity less than the selected animal one.

Compatibility of genotypes in data can also be studied. There is a preliminar genotype compatibility check which compares every animal's genotype with its parents' one.

After this preliminar checking, unknown genotypes can be determined using information provided by animal's progeny. If this check does not obtain animal's genotype, a second one using its parent's genotype could be applied.

References

FALCONER DS., 1989. Introduction to Quantitative Genetics, 3rd ed. Longman. London.

MEUWISSEN THE, LUO Z., 1992. Computing inbreeding coefficients in large populations. Genet Sel Evol., 24:305-313.

BOICHARD D., MAIGNEL L., VERRIER E, 1997. The value of using probabilities of gene origin to measure genetic variability in a population. Genet. Sel. Evol. 29:5-23.

ITEA (2000). Vol. 96A N.° 3, 189-190

EVALUACIÓN BAJO LA HIPÓTESIS DE HERENCIA MIXTA; USO DE INFORMACIÓN DE QTLs EN LA PREDICCIÓN DE VALORES GENÉTICOS POLIGÉNICOS

J.A. Baro*, C.E. Carleos, R. Pong-Wong***, R. Álvarez***

* Unidad de Mejora Genética, Pricipado de Asturias, España

** Departamento de Estadística, Universidad de Oviedo, España

*** Genetics and Biometry Unit, Roslin Institute

Introducción

Para la estimación de valores genéticos aditivos se recurre a las ecuaciones del modelo mixto con datos fenotípicos de los animales y sus parientes. De hecho, es el método de rigor para la identificación de los individuos con el mayor mérito genético en casi todos los programas de mejora genética, y para la monitorización de la respuesta a la selección.

La superación de las asunciones del modelo infinitesimal para incorporar información sobre QTLs en la selección (herencia mixta) requiere la estimación conjunta de los valores de genéticos aditivos individuales y de los alelos del QTL.

Material y métodos

Los efectos de los distintos genotipos de un QTL se estiman a menudo mediante Mínimos Cuadrados Ordinarios (OLS). Cuando algunos de los animales están emparentados, el hecho de que se compartan alelos en loci distintos introduce correlaciones entre los residuales que no se tienen en cuenta en OLS. Similarmente, la estimación de valores genéticos aditivos mediante el modelo mixto e ignorando el genotipo para el QTL puede llevar a estimas sesgadas debido al empleo de asunciones incorrectas sobre el compartimiento de alelos del QTL.

Se presentan resultados para un programa informático (PONG-WONG Y WOOLLIAMS, 1996) que tiene en cuenta la segregación mendeliana de genes de efecto discreto de incidencia conocida y de un fondo poligénico. Dicho procedimiento estima conjuntamente el efecto de los genotipos del QTL y los valores poligénicos aditivos mediante muestreo de Gibbs usando estimas probabilísticas de la matriz de incidencia del QTL y es por tanto aplicable a datos con observaciones faltantes para los genotipos del QTL, lo que es casi inevitable en un programa de mejora genética.

Resultados

Se utilizó un conjunto de más de 2000 registros de varias medidas zoométricas y pesos obtenidos de la estación de testaje de la raza Asturiana de los Valles con 210 animales genotipados para el gen de la hipertrofia muscular (mh). Se obtuvieron estimaciones de heredabilidad del fondo infinitesimal, cuando se incluía el QTL en el modelo

Cuadro 1

	$h^2_{(MIX\,|\,INF)}$	a	d
Peso	.287 | .215	-30.1	2.79^{NS}
Altura	.142 | .257	-2.22	$-.84^{NS}$
Longitud	.354 | .368	-3.06	-2.90^{NS}
Ancas	.502 | .478	2.18	$-.45^{NS}$
Peso	.011 | .002	$-.25^{NS}$	$-.44^{NS}$
Torax	.234 | .157	2.26	-1.16^{NS}
Testículos	.206 | .263	-1.78	$-.94^{NS}$

(NS) no significativamente distinto de cero al nivel 95%.

Cuadro 2

	Peso	Altura	Longitud	Ancas	Pecho	Torax	Testículos
Correlacción	.91	.82	.63	.98	.81	.96	.38

(MIX) o se ignoraba (INF), y para la aditividad y dominancia del QTL.

Se estudiaron los cambios en la correlación de Spearman entre el rango obtenido siguiendo el modelo infinitesimal o el modelo de herencia mixta, cuando se seleccionaban los 50 mejores animales de los genotipados.

Conclusiones

Se encontró que el alelo mh reduce el peso y el tamaño corporal en altura y longitud, así como el tamaño de los testículos, pero incrementa el diámetro torácico y la anchura de ancas. Su efecto dominante no es en ningún caso significativo. Dado el pequeño número de conexiones existentes en el pedigree, no se pudieron extraer conclusiones definitivas sobre la partición de la varianza genética, aunque la segregación del alelo parece ser responsable de toda al variablidad observada en el tamaño testicular

Se observaron diferencias notables para algunas variables en el ranking de los animales cuando se ordenaban por el valor obtenido aplicando un modelo que incluyese o no el efecto del gen mayor.

Agradecimientos

Este trabajo se desarrolló en el marco del proyecto FEDER 1fd97-0042. Agradecemos a la Asociación de ganaderos ASEAVA su colaboración.

Bibliografía

HOFER A, KENNEDY B.W., 1993. Genetic evaluation for a quantitative trait controlled by polygenes and a major locus with genotipes not or only partially known, Genet. Sel. Evol. 25, 537-555.

PONG-WONG, R., WOOLLIAMS, J.A., 1996. Estimating major gene effects with partial information using Gibbs sampling, TAG 93:1090-1097.

ITEA (2000), Vol. 96A N.° 3, 191-198

LA HIPERTROFIA MUSCULAR HEREDITARIA: GÉNESIS DE ALELOS DOMINANTES NEGATIVOS DE LA MIOSTATINA (GDF-8) MURINA

A. Barroso, L.J. Royo, J. Cañón, S. Dunner

Laboratorio de Genética, Departamento de Producción Animal, Facultad de Veterinaria, Avda. Puerta de Hierro s/n, 28040 Madrid, España

RESUMEN

La miostatina es un inhibidor extracelular del desarrollo muscular. Pertenece a la familia de factores de crecimiento TGF-β, y con ellos comparte una serie de características estructurales y funcionales. Basándonos en estas características diseñamos unas construcciones génicas con la intención de crear una serie de alelos dominantes negativos de la miostatina murina, como modelo para la obtención, en un futuro, de terneros que manifiesten la hipertrofia muscular en primera generación por cruzamiento de cualquier hembra con un macho modificado portador del alelo dominante negativo.

Palabras clave: TGF-β, Miostatina, Hipertrofia muscular, Gen dominante negativo.

SUMMARY

Myostatin is an extracellular negative regulator of muscle growth. It is a member of the growth factor family TGF-β, having similar structural and functional characteristics. Based on these structural and functional characteristics we have designed several gene constructions in order to create a series of dominant negative alleles of murine myostatin. Our aim is to obtain double-muscled calves in the first generation by mating a modified sire carrying a dominant negative allele with non-modified cows.

Key words: TGF-β, Myostatin, Muscular hypertrophy, Dominant negative gene.

Introducción

Bases moleculares de la Hipertrofia Muscular Hereditaria

La Hipertrofia Muscular (HM) Hereditaria es un síndrome descrito en las especies murina y bovina (donde también es conocido como "doble grupa" o "cularidad") que provoca, como característica más sobresaliente, un incremento generalizado de la masa muscular del animal (figura 1). A nivel macroscópico, se advierte además una reducción muy marcada del tamaño de los órganos, así como un descenso moderado en el índice de transformación.

A nivel microscópico o histológico, el tejido muscular de estos animales se caracteriza



(a) (b)

Figura 1. (a) Toro culón de la raza Asturiana de los Valles. (b) Ratón hipertrófico –a la derecha–
Figure 1. (a) A fullblood Asturiana de los Valles double-muscled bull. (b) Hypertrophic mouse –rigth–

por un incremento en el número, que no en el tamaño, de las fibras musculares, fenómeno conocido como *hiperplasia muscular* (hipertrofia, por tanto, es inexacto). Esta mayor densidad de músculo se compensa con descensos concomitantes en los tejidos conjuntivo y, sobre todo, graso.

En el ganado bovino, este fenómeno ha sido descrito en diversas razas europeas, tanto belgas (Blanco Azul) como italianas (Piamontesa, Marchigiana), francesas (Charolés, Limousine) o españolas (Asturiana de los Valles, Pirenaica).

El esfuerzo investigador de los últimos cuatro años ha permitido descubrir el gen responsable de la HM, el gen de la miostatina (MSTN) o GDF-8 (*Growth and Differentiation Factor* - 8) (McPherron *et al.*, 1997), que está constituido por tres exones y dos intrones (figura 2). La proteína que codifica, encuadrada en la superfamilia de los TGFβs (*Transforming Growth Factors*), actúa como regulador extracelular negativo del crecimiento muscular (Lee & McPherron, 1999). Está demostrado que las mutaciones que dan lugar a una MSTN inactiva –como la *deleción11* bovina (mutación n.º 5 de la figura 2) (Grobet *et al.*, 1997) o la *Cmpt* murina (Szabó *et al.*, 1998)– son responsables del crecimiento muscular exagerado que caracteriza este síndrome.

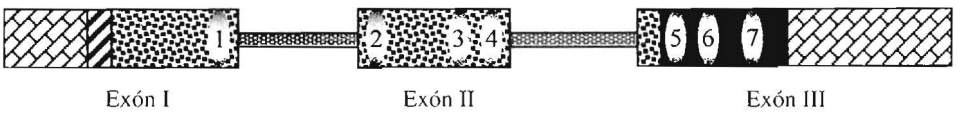


Figura 2. Estructura del gen de la MSTN murino y bovino. Se numeran las mutaciones bovinas naturales (1: F94L; 2: nt419 (del7-ins10); 3: Q204X; 4: E226X; 5: nt821(del11); 6: E291X; 7: C313Y). 5' y 3'-UTRs (UnTranslated Regions) Péptido señal Prodominio Intrones I y II Péptido bioactivo
Figure 2. Structure of murine and bovine MSTN gene. Naturally occurring mutations in bovine are numbered

En la raza bovina Asturiana de los Valles, las canales culonas, con una mayor proporción de carne de primera, alcanzan un precio de mercado muy superior al de las normales. Por ello, es muy interesante intentar conseguir culones en primera generación como resultado del cruce entre un animal culón y uno normal. O, lo que es lo mismo, culones heterocigotos, útiles tanto en el caso de cruzamientos intrarraciales, en razas de carne (como la Asturiana de los Valles); como en el caso de cruzamientos industriales, destinados a aumentar el valor residual de los terneros de razas lecheras (como la Frisona) que no vayan a reposición.

Sin embargo, todas las mutaciones descritas hasta el momento en la MSTN bovina (2-7 en la figura 2) son recesivas, por lo que sólo los homocigotos mutantes manifiestan hipertrofia muscular. Por consiguiente, la consecución de un animal culón en heterocigosis implica la génesis de un alelo dominante negativo, de tal forma que los heterodímeros MSTN normal-MSTN mutante puedan expresarse en el fenotipo. El hecho de que existan un buen número de ejemplos de alelos dominantes negativos naturales en la ruta metabólica del TGFβ1 (WIESER *et al.*, 1993, 1995), así como la dominancia parcial en machos que exhibe la mutación *Cmpt* de la MSTN murina (SZABÓ *et al.*, 1998), refuerzan la hipótesis sobre la viabilidad de alelos dominantes negativos de la miostatina murina y bovina.

Metodología

Estrategias de modificación génica

La génesis de alelos dominantes negativos "artificiales" de la MSTN conlleva definir las mutaciones a producir en la proteína y poner en marcha una estrategia de modificación génica. Esta, a su vez, presupone conocer la estructura del gen, disponer de unos modelos celular y animal asequibles y, finalmente, desentrañar el mecanismo de acción de la proteína a alterar.

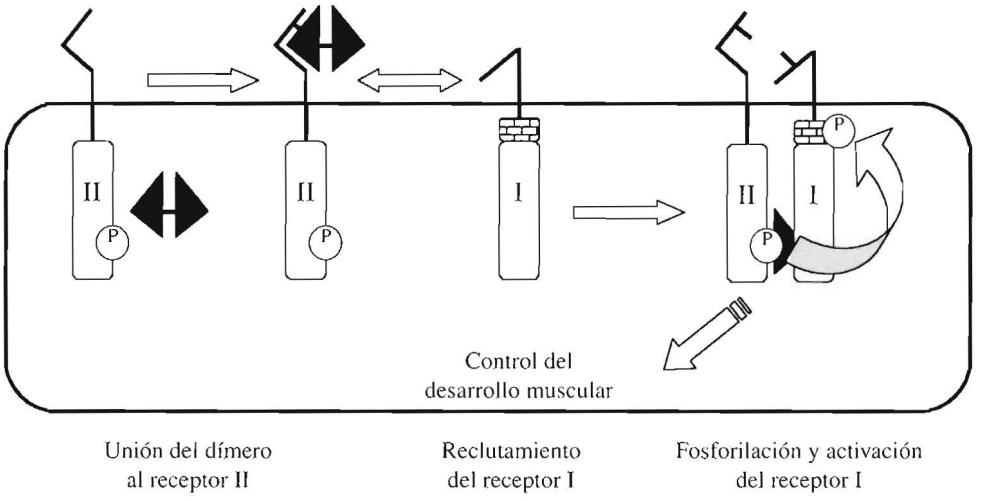


Figura 3. Mecanismo de acción del dímero de miostatina
Figure 3. Initiation of signalling by the myostatin dimer

La estructura del gen está enteramente resuelta desde el clonado y secuenciación del de las especies bovina y murina, fruto de una colaboración con el grupo del Dr. Georges de la Universidad de Lieja (ROYO *et al.*, manuscrito en preparación). Existen asimismo líneas comerciales de ratones y de mioblastos murinos para llevar a cabo los pertinentes análisis funcionales de los dominantes negativos (pues la elevada prolificidad, el reducido intervalo entre partos y su bajo coste hacen del ratón el modelo animal idóneo para este tipo de experimentos). Finalmente, por analogía con TGFβ1, se tiene una visión global bastante certera sobre el modo de acción de la miostatina sobre el tejido muscular (figura 3).

Las diferentes estrategias de modificación génica se pueden dividir con arreglo a tres grandes criterios que se complementan mutuamente (cuadro 1).

Independientemente del tipo de estrategia adoptado, que se detallará más adelante, es necesario aislar o sintetizar in vitro entre 5.000 y 10.000 pares de bases (pb) del gen, dentro de las cuales van incluidas las mutaciones que nos interesa provocar, y clonarlas en un plásmido: el conjunto se conoce como construcción génica. El objetivo final es sustituir la copia normal del gen de la MSTN del genoma de las células por la modificada, fenómeno conocido como Recombinación Homóloga (RH).

Construcciones dominantes negativas de la miostatina murina

En este trabajo, se han diseñado cuatro construcciones dominantes negativas de la miostatina murina, combinando las estrategias de los apartados A.2, B.1 y C.1 de la cla-

Cuadro 1

A. Funcionalidad del gen
A.1. Estrategias *Knock-**Out***: se inactiva completamente el gen bajo estudio eliminando físicamente una parte fundamental del mismo. Esta estrategia no permite obtener dominantes negativos de miostatina, pues queremos que el gen modificado produzca una proteína completa, si bien afuncional, que secuestre la proteína normal y ejerza de esa manera la dominancia negativa. A.2. Estrategias *Knock-**In***: se introducen pequeños cambios en el gen, que nos permiten modular su expresión y/o el mecanismo de acción de la proteína que codifica. Este abordaje sí nos permite obtener proteínas dominantes negativas.
B. Control de la expresión génica
B.1. *Constitutiva:* no se modifica el control endógeno de la expresión del gen. B.2. *Condicional:* se añaden elementos exógenos al gen, que nos permiten controlar a voluntad su expresión de una forma cuantitativa, temporal o espacial.
C. Fase de síntesis
C.1. Técnicas de *Gene Targeting*: se actúa directamente sobre el ADN genómico. C.2. Técnicas *Antisense*: inciden sobre el ARN mensajero. C.3. Mutaciones dominantes negativas: actúan sobre la proteína.

sificación general antes mencionada. Cada una de ellas ha sido sintetizada de dos formas: como *Knock-In* (para ser integrada de forma estable por RH) y como construcción de interacción (para ser expresada en un sistema eucariota de transcripción-traducción).

Resultados y discusión

Mutagénesis dirigida por PCR

Los *Knock-In* se han sintetizado simultaneando técnicas de *Long*-PCR (BARNES, 1994; CHENG *et al.*, 1994) con técnicas de mutagénesis dirigida por *overlap*-PCR (Ho *et al.*, 1989) (figura 4). De esta manera, se han amplificado unos 8.000 pb del gen de la MSTN con una serie de mutaciones (cuadro 1) que inciden, respectivamente, en:

- Las cisteínas (C^1 y C^2) de interacción con los receptores de membrana: construcción de las Cisteínas.
- La diana de proteolisis o *Proteolytic Site* que separa el prodominio del péptido bioactivo: construcción del *Proteolytic Site*.
- El codón STOP de la proteína: construcción de los STOPs.
- La cisteína que forma el puente intercatenario entre dos monómeros de MSTN: construcción de la Cisteína Intercatenaria.

Las construcciones de interacción se han sintetizado siguiendo un procedimiento de overlap-PCR análogo al de los Knock-In (PCR 2-PCR 3 de la figura 4). En estos casos, sin embargo, sólo es necesario generar la secuencia codificante del gen, de pequeño tamaño (cuadro 3), por lo que se han llevado a cabo por PCR clásica.

Knock-In constitutivos

El objetivo final de las construcciones diseñadas es modular la función, que no la expresión, de la MSTN endógena alterando

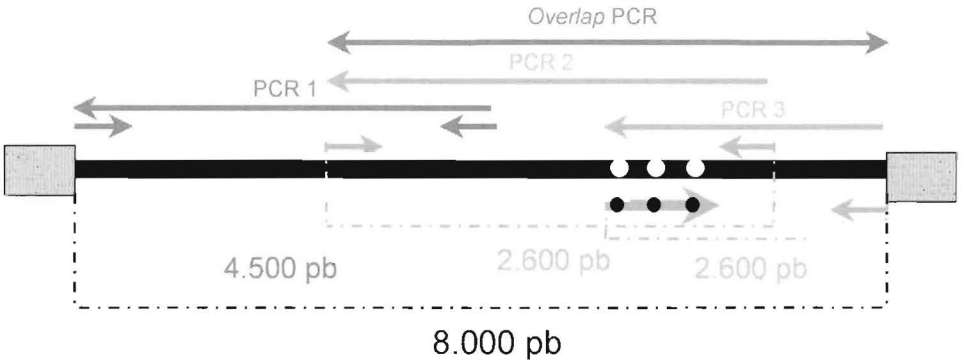


Figura 4. Estrategia molecular común de los cuatro *Knock-In* de la MSTN murina. PCR 3 es específica de cada construcción. Las regiones solapantes entre PCR 1-PCR2 y PCR 2-PCR 3 permiten empalmar unas con otras. ⊟ Extremos del plásmido ■ Secuencia clonada del gen de la MSTN ● Mutaciones introducidas con el *primer forward* de cada PCR 3

Figure 4. Common molecular strategy to generate the murine MSTN Knock-Ins. PCR 3 is specific for each construction (see table 1). Overlappping regions among PCR 1-PCR 2 and PCR 2-PCR 3 make ligation possible

la interacción de la proteína con el receptor de membrana con unas pocas mutaciones en sitios muy precisos (cuadro 2), que no alteran la estructura terciaria de los monómeros mutantes. No se elimina, pues, parte del gen, por lo que hablamos de *Knock-In*, ni se introducen promotores exógenos para el control de la expresión, por lo que se trata de estrategias constitutivas; se incide, finalmente, directamente sobre el ADN genómico, por lo que se ubican dentro de las técnicas de *Gene Targeting*.

Todas las construcciones se han clonado en *pBluescript* [Stratagene] y crecido en bacterias *AG1* [Stratagene] en dos partes: la región común (PCR 1) y las regiones overlap (PCR 2 + PCR 3) específicas de cada construcción, ambas de unos 4.500 pb (cuadro 2). Combinando la primera con cada una de las otras, se han generado los *Knock-In* completos. Resta tan sólo subclonar la construcción completa en un plásmido de Gene *Targeting* para completar la génesis de los *Knock-In* dominantes negativos.

Construcciones de interacción

El objetivo de esta serie de construcciones es proporcionar una idea del grado de afinidad de cada una de las miostatinas mutantes por la molécula normal, como paso previo a los arduos experimentos de *Gene Targeting*. Para ello, se ha clonado la región codificante de cada uno de los *Knock-In* anteriores (cuadro 3) en un plásmido de expresión eucariota, que nos permitirá sintetizar *in vitro* las miostatinas normal y mutante en un lisado de células de mamífero.

En concreto, la región codificante de la miostatina normal ha sido generada por PCR convencional y clonada en *pET-15b* [Novagen]. Este plásmido de expresión permite sintetizar una MSTN con una pequeña "etiqueta" (*tag*) de histidinas, para la que existen anticuerpos comerciales específicos [Pharmacia] que hacen posible su detección inmunológica.

Cuadro 2
Tamaños de los productos de *Long*-PCR de los *Knock-In* murinos
Table 2
Sizes of the murine Knock-In Long-PCR-Amplified products

Construcción	Aminoácidos mutados	Enzima[1]	PCR 1	PCR 2	PCR 3	Total
Cisteínas	C → V; C → Y	*Aat*II			2663 pb	
Proteolytic Site	S → A; R → P; R → I	*Apa*I	4566 pb[2]	2685 pb[3]	2691 pb	7978
STOPs	STOP → W	*Hind*III			2364 pb	
C^{340}	Inserción de 22 aminoácidos	*Apa*I		2057 pb	2434 pb	8033[4]

(1) Estas enzimas sirven para diagnosticar de una forma sencilla si las mutaciones han sido o no incorporadas.
(2) La PCR 1 y la PCR 2 solapan en 1068 pb.
(3) La PCR 2 y la PCR 3 solapan, respectivamente, en 868, 896 y 569 pb en las tres primeras construcciones.
(4) El tamaño final es ligeramente superior por la inserción de 66 pb (22 aminoácidos).

Cuadro 3
Tamaños de los productos de *Overlap*-PCR de las construcciones de interacción de murinas
Table 3
Overlap-PCR sizes of the murine interaction constructs

Construcción	PCR 1[5]	PCR 2	Total
Cisteínas	395 pb	492 pb	527 pb
Proteolytic Site	1115 pb[6]	523 pb	1247 pb
STOPs	395 pb	196 pb	527 pb
Cisteína Intercatenaria	272 pb	266 pb	593 pb

(5) La PCR 1 y la PCR 2 solapan, respectivamente, en 360, 391 y 64 pb en las tres primeras construcciones.
(6) Producto generado por RT-PCR para abarcar la secuencia completa del prodominio (ver figura 2).

Cada una de las regiones codificantes de las miostatinas se han generado por *overlap*-PCR (fusión de las PCR 1 y 2 del cuadro 3), introduciendo las mismas mutaciones de los *Knock-In* (cuadro 2), y se han clonado en el plásmido de expresión *pSG5* [Stratagene]. Los análisis de interacción se harán combinando dos a dos cada una de las construcciones mutantes con la MSTN normal en un sistema TNT de transcripción-traducción [Promega].

Agradecimientos

Este trabajo ha sido financiado por ASTURGEN S.L.

Bibliografía

BARNES W.M., 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from l bacteriophage templates. Proceedings of the National Academy of Sciences USA, 91, 2216-2220.

CAPPUCCIO I., MARCHITELLI C., SERRACCHIOLI A., NARDONE A., FILIPPINI F., AJMONE-MARSAN P., VALENTINI A., 1998. A G-T transversion introduces a STOP codon at the mh locus in hypertrophic Marchigiana beef subjects. Animal Genetics, 29 (Suppl. 1), 51.

CHENG S., FOCKLER C., BARNES W.M., HIGUCHI R., 1994. Effective amplification of long targets from cloned inserts and human genomic DNA. Proceedings of the National Academy of Sciences USA, 91, 5695-5699.

GROBET L., PONCELET D, ROYO L.J., BROUWERS B., PIROTIIN D., MICHAUX C., MÉNISSIER F., ZANOTTI M., DUNNER S., GEORGES M., 1998. Molecular definition of an allelic series of mutations disrupting the myostatin function and causing double-muscling in cattle. Mammalian Genome, 9, 210-213.

GROBET L., ROYO L.J., PONCELET D., PIROTIIN D., BROUWERS B., RIQUET J., SCHOEBERLEIN A., DUNNER S., MÉNISSIER F., MASSABANDA J., FRIES R., GEORGES M., 1997. A deletion in the myostatin gene causes double-muscling in cattle. Nature Genetics, 17, 71-74.

HO S.N., HUNT H.D., HORTON R.M., PULLEN J.K., PEASE L.R., 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene, 77, 51-59.

LEE S.J., MCPHERRRON A.C., 1999. Myostatin and the control of skeletal muscle mass. Current Opinion in Genetics and Development, 9, 604-607.

McPherrron A.C., Lawler A.M., Lee S.J., 1997. Regulation of the skeletal muscle mass in mice by a new TGFb superfamily member. Nature, 387, 83-90.

Szabó G., Dallmann G., Müller G., Patthy L., Soller M., Varga L., 1998. A deletion in the myostatin gene causes the compact (Cmpt) hypermuscular mutation in mice. Mammalian Genome, 9, 671-672.

Wieser R., Attisano L., Wrana J.L., Massagué J., 1993. Signaling activity of transforming growth factor b type II receptors lacking specific domains in the cytoplasmic regions. Molecular and Cellular Biology, 13, 7239-7247.

Wieser R., Wrana J.L., Massagué J., 1995. GS domain mutations that constituvely activate TbR-I, the downstream signaling component in the TGFb receptor complex. EMBO Journal, 14, 2199-2108.

Wrana J.L., Attisano L., Wieser R., Ventura F., Massagué J., 1994. Mechanism of activation of the TGF-b receptor. Nature, 370, 341-347.

ITEA (2000), Vol. 96A N.° 3, 199-219

LA PERSPECTIVA BAYESIANA. TRES NUEVOS DIÁLOGOS ENTRE HYLAS Y FILONÚS

A. Blasco

Departamento de Ciencia Animal, Universidad Politécnica de Valencia, España

Nota preliminar

Los diálogos entre Hylas y Filonús fueron descubiertos por el obispo Berkeley (a quien una famosa Universidad debe su nombre) a principios del siglo XVIII. Berkeley quedó tan impresionado por las argumentaciones de Filonús (sin duda un seudónimo, como luego aclararé), que basó su filosofía en los razonamientos encontrados en los diálogos. Su libro *Tratado sobre los Principios del Conocimiento Humano* no es otra cosa que un desarrollo de los motivos que expone Filonús para dudar de que se pueda llegar a un conocimiento real del mundo, e incluso de que exista otro mundo distinto del de las impresiones sensibles que recibimos. Durante muchos años ha sido una tarea favorita entre los filósofos de todas las escuelas el rebatir a Filonús, pero las argumentaciones de este protagonista de los diálogos son mucho más difíciles de refutar de lo que parece a primera vista. Curiosamente una refutación a Filonús la ofrece el propio Hylas, y se encuentra recogida en otros tres diálogos que exhumó el conocido filósofo de la ciencia Mario Bunge, si bien ciertos motivos que expondré más adelante me hacen sospechar que son en parte apócrifos. Andaba yo ocupado en discernir qué parte de los diálogos publicados por Bunge podían ser considerados auténtica transcripción de las conversaciones entre Hylas y Filonús, y qué parte debía desecharse como añadida por copistas posteriores, cuando di con una serie de apuntes de caligrafías diversas, que reconstruían otros diálogos que tuvieron lugar por la misma época. La dificultad de la reconstrucción radica en que los fragmentos recuperados no llegan en orden, por lo que lo que yo propongo aquí es el que me parece más lógico para seguir las argumentaciones, aunque queda el campo abierto a otros investigadores para proponer no sólo un orden distinto, sino para añadir algunos párrafos a los que yo no encontré lugar, o a interpretar las ausencias e hilarlas mejor con los discursos conservados.

En el primer diálogo Filonús plantea el problema de la inducción a partir de un argumento de Hylas (contenido en el tercer diálogo recogido por Bunge), que le parece poco convincente. En opinión de Hylas la inducción es posible porque si no fuera así no sería tampoco posible la ciencia ni la técnica. Filonús le hace ver que la seguridad en las predicciones científicas es imposible, lo que Hylas admite de mal grado, aunque contrarresta la argumentación de Filonús aduciendo que al menos podemos conocer las cosas de forma probable. Aquí sobreviene una incursión dialéctica sobre lo que podemos entender por probabilidad, y Filonús, cuyo nombre significa "ligado a la mente"[1],

1. Lo que obviamente hace pensar que tanto su nombre como el de Hylas son meros seudónimos.

le hace ver a Hylas que sólo considerando la probabilidad como un grado de creencia personal es posible la inducción. Esto sorprende a Hylas (cuyo nombre quiere decir "ligado a la materia), a quien le parece que un grado de creencia personal es inadmisible para inferir algo de la Naturaleza, pero a lo largo del diálogo va comprendiendo los inconvenientes de considerar la probabilidad de manera externa al investigador. Finalmente acepta que la probabilidad sea definida en sentido amplio como este estado de creencias, pero marcha preocupado por las consecuencias que tiene para la ciencia y la interpretación del mundo la transformación de una herramienta que él creía objetiva en un medidor de la incertidumbre de tipo subjetivo. En el segundo diálogo Hylas confiesa a su amigo Filonús su desazón por no poder huir del subjetivismo al ocuparse de algo tan aparentemente objetivo como la actividad científica, y en particular por la necesidad de recurrir a opiniones previas sobre los resultados que se van a obtener, expresadas a menudo en forma vaga, pero Filonús le hace ver que precisamente el uso de esa información previa, aunque sea irrelevante, permite usar la probabilidad para precisar la incertidumbre asociada a los experimentos, lo que no es posible usando otras teorías. Cuando parece que todo haya concluido, resuelto el problema de la inducción, en el tercer diálogo Hylas objeta a Filonús que los métodos de creencias acaban siempre por asumir indirectamente hipótesis no contrastadas. Filonús repara en que las razones de Hylas son válidas, pero también en que, de admitirlas, aunque la actividad científica todavía sería posible sobre bases racionales, quedaría sin resolver el problema de la inducción. Ambos parten preocupados, prometiendo reunirse para continuar indagando sobre la cuestión.

He agrupado arbitrariamente los diálogos en tres, porque así lo hizo Berkeley y así lo repitió Bunge, pero responden más que a un orden temático que al desarrollo de la discusión tal y como infiero que se produjo. El lector que quiera conocer mejor los argumentos de los protagonistas, puede recurrir a las ediciones de los diálogos previos que se encuentran en las editoriales Alianza y Ariel, recuperados por Berkeley y Bunge respectivamente. En algún momento he tenido que rellenar los huecos encontrados en los manuscritos, anotando el tema que se introducía en la discusión, aunque han sido pocas las veces que me visto obligado a interrumpir la conversación. Las notas a pie de página son enteramente mías.

Primer diálogo

– *Hylas.* ¡Buenos días, Filonús! ¿Cuál ha sido el resultado de tus meditaciones? ¿Admites ya que podemos conocer más cosas de las que los sentidos nos muestran? ¿O por el contrario aún piensas que el mundo material nos es desconocido?

– *Filonús.* Lo admito, Hylas, y admito que podemos concluir que hubo hechos hace muchos años de los que nadie fue consciente, que podemos construir entes como los números irracionales, que nadie ha percibido[2], y que tenemos información innata[3] que no procede de nuestras percepciones. Pero

2. Estos argumentos se recoge en el primero y segundo diálogo de los publicados por M. Bunge, respectivamente.

3. Este argumento no figura en ninguno de los diálogos recogidos. Pertenece probablemente a un fragmento perdido.

usaste ayer[4], querido amigo, un argumento que me inquietó tanto que no he podido apartarlo de mi cabeza desde que nos separamos. Me dijiste que podíamos inferir una ley de la Naturaleza a partir de la inducción de numerosísimos casos particulares[5].

– *Hylas.* Y así lo creo. ¿Cómo si no funcionarían las ciencias? ¿Cómo podríamos construir puentes sin la incertidumbre de que no fueran a caer, despeñando a todos los que sobre él se encuentran, si nuestra experiencia no nos mostrara las condiciones que los hacen seguros? ¿Podrías vivir en la incertidumbre de si el Sol saldrá mañana?[6].

– *Filonús.* Entiendo, Hylas, las dificultades que negar tu razonamiento entraña, pero no encuentro inconveniente en afirmar que es posible que el Sol haya desaparecido ya y no lo sepamos hasta dentro de los escasos minutos que tarda su luz en llegar hasta nosotros[7]. Mi fe en que el nuevo puente no caerá no ha impedido que cayeran otros, y sólo podría decirte que si vivo sin esas preocupaciones se debe a que mi naturaleza no me hace el pensar en ellas, y actúo como si supiera cierto lo que no sé ni tan siquiera como probable, y no me inquieto pensando en que el Universo pueda consumirse hoy o mañana[8]. No comprendo cuál es la regla que, según tu parecer, Dios ha creado para que al enumerar varios casos particulares podamos inferir sobre los casos que no conocemos.

– *Hylas.* Admito, Filonús, que no podamos hablar con la seguridad de quien tiene perfecto conocimiento de todos los casos que la Naturaleza provee, pero no me negarás que sí podemos calificar ciertos sucesos como de altamente improbables.

– *Filonús.* No entiendo, Hylas, qué quieres decir con la palabra "improbable". ¿Me hablas del número de puentes que caen respecto a los que permanecen firmes?

– *Hylas.* Resulta chocante tu propuesta, y la creo animada por el juego de la ironía más que por el de la discusión que intenta esclarecer lo oscuro de estos conceptos; pero aún así podría admitir lo que dices como una expresión de la incertidumbre acerca de que un puente se desplome.

– *Filonús.* ¿Y cuántas veces estalló el Sol para que podamos calcular la probabilidad de que vuelva a esfumarse?

– *Hylas.* Bromeas, Filonús, y te ruego que centremos nuestra conversación en ejemplos concretos sobre los que te pueda dar respuesta.

– *Filonús.* Con tu forma de razonar, Hylas, pocos son los ejemplos de incertidumbre a los que podríamos asociar proba-

4. Tómese "ayer" no en sentido que se le da hoy en día, sino en su sentido etimológico: *ad hieri,* hace algún tiempo, en tiempo pasado.

5. Ver el primer diálogo de los recogidos por Bunge.

6. Este ejemplo es habitualmente asociado a HUME (1739), quien publicó su Tratado sobre la Naturaleza Humana en fecha posterior a la ocurrencia de las conversaciones entre Hylas y Filonús, pero se encuentra ya en Sexto Empirico, en el siglo II de nuestra era (ver, p. ej., COPLESTON, 1963).

7. Este fragmento podría muy bien ser apócrifo, puesto que la velocidad de la luz no se calculó hasta el siglo XIX.

8. Recuerda a la "fe animal" de la que en este contexto hablaba el mejor filósofo español, mi tocayo Jorge Agustín Santayana (profesor de filosofía en Harvard a finales del siglo XIX y principios del XX).

bilidades. Si hacemos un experimento para inferir tal o cuál cosa, ¿cómo podemos asociar a nuestros resultados lo que es probable, si no vimos nunca en cuántos casos fallaban nuestras predicciones?

– *Hylas.* Puedes imaginar que repetiste ese experimento infinitas veces y que tu resultado es el de una de las posibles repeticiones.

– *Filonús.* Mucha imaginación me pides, amigo Hylas, y te ruego que me pongas un ejemplo de tu proceder, porque me encuentro desorientado y no sé como puedo inferir leyes generales de cosas que no he visto.

– *Hylas.* Pongamos, Filonús, que lanzo una moneda al aire y anoto si salió cara o cruz.

– *Filonús.* Admito el ejemplo, Hylas.

– *Hylas.* Tras muchas repeticiones de este experimento, aproximadamente la mitad de las veces habrá salido cara y la mitad cruz, con lo que puedo decir que la probabilidad de que en un nuevo lanzamiento sea cara el lado del que caiga la moneda es del 50%. Ahí ves cómo infiero sobre cosas que no he visto.

– *Filonús.* Pero si al cabo de mil repeticiones, un 60% de las veces hubiera salido cara y un 40% cruz, ¿qué dirías?

– *Hylas.* Sin duda postularía que la moneda está trucada.

– *Filonús.* ¿Y si arrojara un millón de veces la moneda, y hubiera caído de cara en el 55% de las ocasiones, cayendo en cruz las restantes, qué dirías entonces?

– *Hylas.* También diría que la moneda está trucada.

– *Filonús.* Luego la única definición posible de moneda es "aquél objeto que al ser lanzado al aire muchas veces cae de cara aproximadamente en la mitad de las ocasiones", ¿no es así?

– *Hylas.* Veo Filonús que me estás acusando de introducir lo definido en la definición, pero esto no es así. Lo que yo digo es que con mi moneda trucada, que un 60% de las veces me da cara y un 40% cruz, puedo hacer inferencias sobre qué resultados obtendré en nuevos lanzamientos, y diré entonces que la probabilidad de que me salga cruz el sólo 0,4 si me obstino en seguir utilizándola.

– *Filonús.* Pobre conclusión se me antoja, Hylas, pero me gustaría que me dieras un ejemplo más relacionado con las leyes de la Naturaleza.

– *Hylas.* Pues ya que paseamos por el campo, imaginemos que quisiera conocer cuánto pesan los cerdos Hampshire[9] al nacer. Tomaría un buen número de ellos, los pesaría, e inferiría una ley general de la Naturaleza.

– *Filonús.* ¿Y cómo sabrías que el verdadero valor no es muy distinto del que tú obtuviste?

– *Hylas.* Porque si repitiera un sinnúmero de veces la recogida de lechones y sus pesadas, todos los valores que obtendría se agruparían en torno al valor verdadero.

– *Filonús.* Hylas, me estás pidiendo que concluya algo no sólo de los valores que tú

9. Mi amigo Luis Silió me hace notar que la denominación de los cerdos de esa zona como "Hampshire" no se produjo hasta principios del siglo XX. Puede ser un fragmento apócrifo o un error de transcripción por "Berkshire", cerdos de los que ya hay descripciones desde mediados del siglo XVIII.

obtuviste, sino de los que podrías haber obtenido y no lo hiciste. ¿Te parece un proceder recto el concluir las leyes basándote en algo que no has hecho?

– *Hylas.* Yo no concluyo nada de los experimentos que no hice, sino del mío y de mi conocimiento de qué ocurriría si los repitiera[10].

– *Filonús.* Supongamos, entonces, que has usado una balanza que pesa hasta los diez kilos[11], y supongamos que tus medidas se agrupan en torno a un kilo; digamos que fueron 0,6, 0,8, 1, 1,2 y 1,4 kg.

– *Hylas.* Yo diría entonces que la media del peso es de 1 kg

– *Filonús.* Pero ¿qué dirías tú si yo te dijera que la balanza estaba estropeada, y no podía medir pesos de más de 1,4 kg?

– *Hylas.* Entonces tendría que aumentar mi estimación, por el hecho de que si hubiera repetido muchas veces el experimento, me habrían salido medidas superiores a 1,4 kg que yo no habría podido detectar[12].

– *Filonús.* Pero si a continuación nuestro amigo el obispo Berkeley nos dijera que había arreglado la balanza sin que yo me apercibiera, ¿qué valor asignarías a la media de los pesos al nacer de nuestros Hampshire?

– *Hylas.* Sin duda de nuevo 1 kg.

– *Filonús.* De forma que estás modificando el valor de tu inferencia, no en base a los datos de tu experimento, que nunca superaron el valor a partir del cual la balanza parecía rota, sino en base a los experimentos que podrías haber realizado y no realizaste. ¿Y te parece que un hombre cuerdo debe proceder basando su inferencia en los resultados fantasmas de experimentos que nunca realizó?

– *Hylas.* Hombres de talento he visto yo que así procedían, y a nadie debes culpar si le informas por error invencible de qué resultados puede esperar.

– *Filonús.* Pongamos otro ejemplo: ¿Cómo puedes asociar probabilidades a hechos como el que mañana llueva, que descubramos nuevos mundos, que las empresas en que invertimos nuestro dinero vayan a prosperar, que las decisiones que tomamos en algún momento crucial vayan a ser correctas?

– *Hylas.* No se me ocurre, y no creo que haya hombre en la Tierra que tenga para esto solución.

– *Filonús.* Considera, Hylas, que podamos determinar la incertidumbre expresando nuestro grado de creencia en que se producirá un evento, o nuestra opinión sobre el valor que tomará un carácter como el peso al nacer de los lechones que antes queríamos medir.

– *Hylas.* No entiendo qué me propones, Filonús. ¿Quieres que asigne valores a mis creencias, de forma que sigan las leyes de la probabilidad?

– *Filonús.* Eso pretendo y no otra cosa. Seguro estoy de que antes de medir los lechones no creías tú que fueran a pesar diez kilos, ni tampoco que cien gramos. Asigna,

10. Lo que modernamente diríamos es "de nuestro conocimiento de la distribución de la variable en el espacio muestral".

11. Las medidas de peso van en el original en libras inglesas.

12. Este es un problema estadístico clásico: se trata de encontrar una estima insesgada de datos censurados. La paradoja está explicada formalmente y con detalle en varios textos; por ejemplo, BERGER y WOLPERT, (1982).

pues, a tus creencias un valor, y veamos luego[13] los resultados del experimento.

– *Hylas.* Esto que me propones me parece irregular, Filonús. Yo puedo tener unas creencias y tú otras, y por mucho que mezclemos con ellas los resultados que obtengamos de un experimento, siempre acabarán por ser nuestras creencias personales, nada sólido en lo que basar una inferencia.

– *Filonús.* Quiero hacerte notar, Hylas, que no son los experimentos tan concluyentes, habitualmente al menos, como para confirmar o refutar una teoría, o incluso en muchas ocasiones como para poder precisar bien una medida. Y también te recuerdo que todo resultado está sometido a discusión y a interpretaciones personales.

– *Hylas.* Pero al menos podemos separar lo que dicen los datos de la parte que nosotros añadimos[14].

– *Filonús.* ¿Y qué actitud tomarías tú si los datos te dijeran que los lechones de tu experimento pesaron treinta kilos al nacer? ¿No dudarías al menos de la precisión de tu balanza?

– *Hylas.* Tu pregunta lleva implícita la contestación, pero no sé en qué modifica mi argumento.

– *Filonús.* Y si comentaras con los granjeros de la región el peso de tus lechones, y ellos te dijeran que pesan mucho menos de lo que se ha observado en otras ocasiones, ¿qué dirías?

– *Hylas.* No tengo una respuesta puesto que no hice el experimento, pero supongo que concluiría que los cerdos que medí estaban enfermos, o llegaría a alguna conclusión similar que me permitiera explicar la anomalía de mis datos.

– *Filonús.* ¿Y si te dijera que la información que te han dado algunos de los granjeros con que hablaste no es de fiar, porque todos son unos bellacos y gustan de reírse de los hombres de ciudad; o porque son unos porros y no distinguen una balanza mejor que lo hacen sus propios cerdos?

– *Hylas.* Es obvio que entonces daría poco crédito a las afirmaciones de éstos, y no así a las de los granjeros que fueran de fiar.

– *Filonús.* ¿Y no estarías entonces, mi buen Hylas, concluyendo a partir de informaciones anteriores, evaluadas subjetivamente, y no a partir de tus datos?

– *Hylas.* Así lo estaría haciendo, pero insisto aún en que separaré entonces lo que opino de lo que aporto; quiero decir, los datos que obtengo, de mi interpretación personal sobre ellos.

– *Filonús.* No entiendo bien de qué te sirve tal separación, si no vas a usarla luego. Pero volvamos a mi argumento original: si quieres saber qué tan probables son tus resultados, debes usar alguna forma de evaluación subjetiva de esta probabilidad.

– *Hylas.* Ya te he dicho antes qué entiendo por probabilidad, y cómo evaluaría la incertidumbre de mi experimento, y no he necesitado recurrir a mi mundo personal

13. Me he permitido traducir de forma arcaizante "luego" en el sentido de "inmediatamente", como era el significado de esta palabra en castellano hasta bien entrado el siglo XIX.

14. Aquí nos encontramos con un precedente de la conocida aspiración a "dejar que los datos hablen por sí mismos".

para ello, sino que te he hablado de medidas objetivas.

– *Filonús.* Veo que hicieron poca mella las críticas que hice a tu método. Pero qué dirías si concluyera que tu probabilidad es un caso particular de ésta más amplia que te propongo?

Aquí se produce una cesura que no he podido reconstruir. Es obvio que Filonús le propone a Hylas algo similar al concepto moderno de "intercambiabilidad" de los datos. Cuando el orden en que aparecen los datos en una muestra no afecta a la probabilidad de la misma, los datos son "intercambiables", y en ese caso la teoría de la probabilidad como frecuencia es un caso particular de la teoría subjetiva[15]. Esto afecta a los argumentos de Filonús sólo de forma marginal, pero debe indicarse para conservar la ilación del diálogo.

– *Hylas.* Me preocupa, Filonús, cuanto me dices, y debes dejar que reflexione sobre ello, puesto que es la primera vez que me enfrento al hecho sorprendente de que la probabilidad pueda ser mi estado de opinión y no algo ajeno a mí como hasta ahora creía.

– *Filonús.* Más sorprendente me parecen a mí tus repeticiones mentales de un experimento. Y qué decir de tus probabilidades cuando las aplico a la astronomía o a los sucesos que nos deparará el mañana.

– *Hylas.* Pero es cierto que admitimos con más comodidad las dificultades de una teoría que nos enseñaron en nuestra juventud, y que convivimos con ellas con la misma facilidad que con los problemas que te ocasionan los parientes, considerándolas producto inevitable del destino y aceptándolas con resignación y paciencia, mientras que somos más intolerantes con las dificultades que ocasionan teorías nuevas, por más que vayan a mejorar nuestras vidas y nuestro entendimiento de las leyes que gobiernan la Naturaleza.

– *Filonús.* Es justo cuanto dices, Hylas, y creo que es bueno que dejemos aquí la discusión para que puedas ir acomodando el entendimiento a los nuevos conceptos con los que hoy lo has desafiado. La razón no es, como nos quieren persuadir los filósofos, una máquina perfecta que admite cuanto de nuevo le dan si el silogismo es correcto, y yo he visto a muchos hombres sabios andar empecinados, sin atender a buenas razones, hasta que al poco pudieron cambiar de forma de pensar; y aún así me refiero sólo al caso en que lo hicieron.

Segundo diálogo

– *Filonús.* Buenos días Hylas, ¿cuál ha sido el resultado de tu reflexión. Te dejé ayer preocupado y no quise molestarte durante la tarde, absorto, como estabas, en tus meditaciones. ¿Encuentras hoy a la Naturaleza más inteligible?

– *Hylas.* ¡No me digas más, amigo, que ando enfermo pensando que nada real puedo inferir del Mundo, y que cuanto tengo son las opiniones de las personas con las que he hablado! Yo creía que, si bien no con la seguridad a la que todo hombre aspira, al menos podría afirmar que existen las cosas que yo observo y que están fuera de

15. Quiero decir de una de las versiones de la teoría subjetiva (de la versión "objetiva", si se me permite un juego de palabras). El concepto fue propuesto por Bruno de Finetti, y aunque se encuentra particularmente desarrollado en su Teoría de la Probabilidad (De Finetti 1974), puede encontrarse en cualquier libro de texto bayesiano (p. ej. Press, 1982).

mí. Creía que, mediante el experimento y la pericia, podría deducir las reglas por las que se rige el Mundo, y que así comprendería mejor la obra del Creador y sus fines y sus medios. Ahora me encuentro consternado, puesto que sólo puedo contar con las creencias de mis vecinos, sin que nada externo a cada hombre me explique cuáles son las causas de las cosas.

– *Filonús.* Hylas, me fatiga hablar contigo (y no te molestes, te lo ruego, que sabes que te hablo con amor de amigo), porque vuelves una y otra vez a dudar de la realidad que te rodea[16]. ¿Crees por ventura que tus vecinos son dementes y que van a decirte cosas sin sentido no más se encuentren contigo? ¿No ves que una cosa es que la probabilidad sea subjetiva y otra que no sea razonable?

– *Hylas.* Pero yo antes creía en la realidad de las cosas que tenía que estimar, y sabía que existía un valor verdadero e inmutable, cerca del cual se agruparían los datos que obtuviera en mis experimentos. Ahora se ha esfumado ese valor, ya no existe, ya no lo menciono, mis afirmaciones son sobre la probabilidad de que se obtenga esta o esa cantidad, y el valor verdadero ha desaparecido!

– *Filonús.* No ha desaparecido, aunque estás en lo cierto cuando dices que ahora ya no lo mencionas. ¿Y para qué ibas a mencionar algo que no sólo desconoces sino que jamás podrás conocer? ¿Qué utilidad tiene referirse a un valor no sólo desconocido sino incognoscible? Ese valor existe, sí, y continúa existiendo cuando hablas como yo lo hago; y permanece sólido e inmutable en el mismo mundo que el tipo del cerdo Hampshire, el jarrón ideal, el joven y los caballos[17]. Yo hablo simplemente de la probabilidad de que los lechones al nacer alcancen un determinado peso, no de cuál sea ese valor ideal inaprehensible.

– *Hylas.* Pero yo antes también quería hablar de probabilidades y me lo vetaste.

– *Filonús.* Tú no me hablabas de la probabilidad de que el peso al nacimiento de los lechones fuera una u otra, a la luz de tu muestra, sino de qué probabilidad tenía tu muestra de tener éste o aquél valor. Y si he de serte franco, no entiendo a qué viene tanto interés en saber qué probabilidad tiene tu muestra en lugar de preocuparte por la esencia y el meollo del problema; esto es, qué probabilidad tiene el peso al nacimiento del cochino de tomar éste o aquél valor.

Se produce aquí una censura importante en los papeles originales. Obviamente Hylas y Filonús han estado discutiendo sobre alguna forma de lo que hoy conocemos como *probabilidad inversa*, popularizada a través del teorema atribuido a Bayes (1780)[18]. Este teorema es utilizado por una

16. Este "otra vez" se refiere probablemente a las dudas que expresaba Hylas al principio del segundo de los diálogos recogidos por el obispo Berkeley.

17. Esta última afirmación hace referencia a un pasaje del "Fedro", uno de los diálogos de Platón en los que se establece el mundo de las ideas en donde yacen los universales, y del que los objetos del mundo son copias imperfectas. Lo único que no cuadra en el pasaje es el tipo ideal del cerdo Hampshire, por lo que lo tengo por una adición apócrifa.

18. Esta atribución es dudosa, como luego comentaré. Bayes fue un oscuro clérigo que no llegó a publicar ningún trabajo matemático en vida. Era miembro de la Royal Society por unos estudios sobre metafísica. El manuscrito original de su artículo no se conserva y debemos confiar en la adscripción a Bayes que su amigo R. Price hizo al presentar la publicación en la Royal Society.

escuela de estadística como base de la inferencia científica no demostrativa. Dada una hipótesis H y un conjunto de evidencias experimentales (de datos, por ejemplo) y, la inferencia no demostrativa consiste en establecer la probabilidad de esa hipótesis a la luz de la evidencia presentada por el experimento. El teorema establece que esta probabilidad es[19]

$$P(H|y) = P(y|H) \cdot P(H) / P(y)$$

donde P(y|H) es la probabilidad de los datos en caso de que la hipótesis fuera cierta, P(H) es la evidencia previa en favor de la hipótesis antes de hacer el experimento, y P(y) la probabilidad absoluta de que se presenten los datos del experimento y; esto es, teniendo en cuenta tanto esta hipótesis H como las alternativas. Este teorema no es objeto de discusión como tal teorema, puesto que la demostración es sencilla. La discusión se centra, desde entonces, en el significado, estimación y uso de la evidencia previa al experimento P(H).

– *Filonús.* Y te hice notar, Hylas, que la única forma en la que puedes hablar de probabilidad asociada a tus datos es considerando la probabilidad previa, la que se genera de tus observaciones, y finalmente la que resulta de combinar ambas[20]. No de otra manera tus afirmaciones pueden ir unidas a una probabilidad.

– *Hylas.* ¿Pues qué valor tendrán así mis predicciones? ¿No te das cuenta que si doy a mi opinión *a priori* un gran predicamento, el resultado estará determinado por esta opinión que yo he introducido? ¿Qué no ves que así podría yo comunicar a la Royal Society cualquier resultado que me viniera en gana?

– *Filonús.* Yo parto, Hylas, de que eres hombre honrado[21] y de que tu opinión previa no se halla sopesada por tus intereses. Pero de todos modos dime, Hylas, si tan seguro estás de tu opinión previa, ¿para qué quieres realizar el experimento?

– *Hylas.* No sé qué quieres decir.

– *Filonús.* Quiero decir que si tu opinión previa tiene una alta probabilidad, no hay razón para que lleves a cabo un experimento, puesto que sólo conseguirías mermar tu hacienda sin que tu sabiduría aumentara o fuera muy distinta a la que tenías antes de empezar. Sólo cuando tu opinión *a priori* es vacilante es cuando tiene sentido el buscar evidencias en la Naturaleza que corroboren o desmientan las ideas que vagan por tu mente.

– *Hylas.* Puede que yo esté interesado en comprobar algo de lo que ando muy seguro.

– *Filonús.* Te contradices, Hylas; si andas muy seguro de ello no habrá experimento que te quite tal seguridad, y si un experimen-

19. Uso la nomenclatura de Jeffreys (1932), quien expresa la "H dado y" como "H|y". A veces se encuentra en la literatura "H/y", pero el uso de la barra inclinada confunde por su similitud con la operación de dividir.

20. Esta es una expresión del Teorema de Bayes que parece chocante en boca de Filonús, puesto que este teorema fue publicado de forma póstuma en 1780. Sin embargo STIGLER (1983) ha hecho notar que el teorema en cuestión figura ya en un libro publicado en 1750, en el que el autor indica que es préstamo que toma de un amigo. Stigler también evidencia cuán improbable es que el tal amigo fuera el propio Bayes, por lo que debemos atribuir la autoría del teorema a un matemático desconocido. Tal vez el autor de este teorema sea el propio Filonús.

21. "Honest" en el original inglés. Se debe traducir por "honrado" y no por "honesto", puesto que en castellano la honestidad sólo hace referencia a la virtud sexual.

to es capaz de quitártela, tu opinión previa no era tan firme como me quieres persuadir de que lo era.

– *Hylas.* Acepto lo que me dices, pero aún tengo otra objeción. Si sólo debo contrastar mi opinión *a priori* cuando es vaga, ¿qué provecho saco, pues, a esa opinión? ¿Para qué me sirve tener una teoría que une la opinión previa a la de los datos, si luego sólo la debo utilizar cuando la opinión previa carece de valor?

– *Filonús.* Estás exagerando, Hylas. Una cosa es que la opinión previa domine y otra que no pueda combinarse con la que los datos te dan, en lo que vale. De todas formas, si ello te inquieta, yo te aconsejaría que tomaras tantos datos como para que la opinión previa no influyera en tu resultado.

– *Hylas.* Eso que me propones, Filonús, no es consistente con tus planteamientos. Pues, ¿qué ventajas obtendré yo de tu modo de proceder, si al cabo debo ignorar mis opiniones previas? ¿A qué hacer tanta novedad si debo hacer luego lo posible para que mi opinión previa no sea tenida en cuenta? Me hablas primero[22] de la superior coherencia de tu sistema, de su belleza, de su simplicidad; y me propones a continuación que ignore la parte principal en la que está basada su belleza, esto es, la combinación entre probabilidades previas y posteriores al experimento.

– *Filonús.* Yo no te propongo que abandones estas probabilidades previas, sólo te digo que si estás inseguro sobre cómo debes expresarlas y prefieres confiar en la información que dan tus datos, tienes un medio de hacerlo.

– *Hylas.* Te confieso, Filonús, que tu sistema pierde así gran parte de su encanto. No sé qué gano entonces, si al final son los datos quienes me van a sacar de mi ignorancia.

– *Filonús.* Siempre son los datos quienes nos sacan de nuestra ignorancia, sean datos que nosotros obtenemos o los que otros obtuvieron antes que nosotros. Nuestra opinión *a priori* no debe ser parcial, infundada, o movida por el interés, sino antes bien debiera ser tal que varios hombres cuerdos pudieran abrazarla, dentro de la incertidumbre con que necesariamente la debemos expresar.

– *Hylas.* ¿Y no es mejor que sean los datos de tu experimento quienes te saquen de tu ignorancia, y no esa vaga idea de la probabilidad *a priori*, que no parece sino un espíritu intangible y de límites borrosos?

– *Filonús.* Te insisto, Hylas, en que la probabilidad *a priori* es necesaria para construir la inferencia. No puedes decir que escoges la hipótesis más probable sin hablar de ella. Esto no es una perversión de mi sistema, sino una obligación de las leyes de la probabilidad.

– *Hylas.* Puedo escoger la hipótesis que haga más probables a mis datos, así no necesito hablar de probabilidades *a priori.*

– *Filonús.* No sé qué quieres decir con esto.

– *Hylas.* Quiero decir que entre dos o más hipótesis alternativas, escogería aquélla que hiciera más probable el que yo fuera a obtener los datos que de hecho he obtenido[23].

22. Se ha perdido este pasaje en el que Filonús habla de la coherencia de su sistema.

23. Hylas propone decidir en base a lo que hoy llamamos "factor de Bayes".

– *Filonús.* ¿Y qué ventajas le ves a ese procedimiento? ¿Porqué crees que es más razonable escoger la hipótesis que hace más probables los datos que vayas a obtener?

– *Hylas.* Yo lo veo intuitivamente razonable. No me parece más adecuado sostener que era sumamente improbable que aparecieran los datos que de hecho obtuve

– *Filonús.* Hylas tú no estás escogiendo la hipótesis que hace más probable el que hayas obtenido esa muestra, sino la hipótesis que *si fuera cierta* haría que tus datos surgieran con máxima probabilidad, que no es lo mismo.

– *Hylas.* No veo la importancia del matiz.

– *Filonús.* Te pondré un ejemplo. Supone que quieres estimar la estatura de los hombres de Escocia, y para ello tomas la estatura de un único escocés, y éste mide 1,60 metros[24].

– *Hylas.* Escasa me parece la muestra, pero continuemos. Yo diría que la hipótesis que haría que ese dato fuera el más probable, es que la media de los escoceses midiera 1,60. Si esto fuera así, ese dato sería el más probable. Si los escoceses fueran un pueblo de enanos cuya media fuera sólo un metro de estatura, el obtener un escocés de 1,60 metros sería altamente improbable, y no sé porqué debo sostener que obtuve un dato altamente improbable, si yo actué escogiendo a este hombre al azar.

– *Filonús.* Quiero hacerte notar de nuevo que has dicho *si esto fuera así*, por lo que en realidad no sabes si es ese el valor que hace tu dato más probable o no.

– *Hylas.* Soy consciente de ello, pero así y todo, no entiendo porqué debería sostener que la media de los escoceses mide dos metros, o un metro y cincuenta.

– *Filonús.* Hylas, la hipótesis que hace tu único dato el más probable es que TODOS los escoceses miden 1,60 metros. Esta y no otra es la hipótesis que hace más probable a tus datos, pues sostener que haya escoceses que midan una cantidad distinta a 1,60 hace más improbable encontrar a un escocés de esa estatura que si la totalidad de ellos mide esta cantidad.

– *Hylas.* ¿Bromeas acaso, Filonús? ¿Qué no son hombres los escoceses, sino máquinas diseñadas por algún espíritu maligno enemigo de la delicadeza en las artes y la perfección en la música?[25] ¿Cómo quieres que sostenga algo tan absurdo como el que todo escocés mide lo que su vecino?

– *Filonús.* ¿Y cómo sabes que los escoceses son hombres, y que sus estatura debe variar, y que no es de esperar encontrar una uniformidad en la naturaleza que no se encuentra en otros lados? ¿No estarás, por ventura, usando de alguna información *a priori?*

– *Hylas.* Aunque así sea, y obviamente lo es, sólo interviene en mi inferencia haciéndome descartar hipótesis absurdas.

– *Filonús.* Haciéndotelas descartar, o dándoles probabilidad *a priori* nula, que es lo mismo. El ejemplo es extremo, Hylas, pero puede hacerte ver que sólo el escoger la hipótesis que, *de ser cierta*, hace a tus datos más probables, es insuficiente para la inferencia, y sólo cuando esta hipótesis está

24. Traduzco pies por metros, como anteriormente libras por kilos.

25. Se infiere que Hylas es inglés.

ponderada por la probabilidad *a priori* puedes escoger lo que realmente deseas; esto es, la hipótesis más probable dados los datos de que dispones.

– *Hylas.* Cuanto dices está bien para escoger entre hipótesis que compiten entre si, pero una vez ya establecido el marco en el que nos moveremos; esto es, que los escoceses son personas y que la estatura de las personas es variable, y que tan probable es encontrar escoceses que midan más que la media como menos, creo que puedo escoger para la estatura de los escoceses el valor que haga más probable el que yo haya encontrado mis datos y no otros.

– *Filonús.* Quieres decir el valor que, *en el caso de que fuera el verdadero valor de estatura de los escoceses*, haría más probable el que fueras a encontrar los datos que obtuviste[26].

– *Hylas.* Sí, Filonús, quiero decir eso, ya me lo has hecho notar varias veces.

– *Filonús.* Pero entonces te encuentras en la misma situación. Imagínate que el escocés que encontraste mide dos metros de altura. ¿Sería razonable pensar que son un pueblo de gigantes?

– *Hylas.* Es muy improbable que fuera a encintrar un valor así escogiendo un escocés al azar.

– *Filonús.* Pues eso es precisamente lo que te quería hacer notar. Tu sistema tiene el defecto de que sólo hace más probable a la muestra *en el caso de que* el valor verdadero sea el que obtuviste. Para evitar el terrible inicio de tu inferencia, "*en el caso de que*", hace falta ponderar ese "*caso*" por su probabilidad, y sólo así obtenemos una inferencia sensata.

– *Hylas.* Estaría de acuerdo si, en efecto, ponderáramos cada caso por su probabilidad, pero no es este tu modo de proceder, esa probabilidad por la que tú ponderas no es en realidad la probabilidad de cada caso, es sólo tu opinión previa. No quiero que mis prejuicios influencien mi experimento hasta tal extremo. Estoy dispuesto a ignorar de entrada qué tan altos espero que los escoceses sean, y para mí cualquier valor que pueda imaginarse tiene en principio la misma probabilidad.

– *Filonús.* No me parece una manera inteligente de actuar, pero aún así te diría que gracias a esa probabilidad constante que te has marcado *a priori* puedes hablar de cuán probable es el que los escoceses midan o no dos metros de altura por término medio[27].

– *Hylas.* No quiero concluir absurdos, Filonús, yo diría que en este caso un solo escocés no es suficiente para mis propósitos y debo tomar un número lo suficientemente grande como para no depender de mis prejuicios.

– *Filonús.* Pues por eso te decía antes que si tanto te preocupa no depender de tus opiniones previas, tomaras pues tantos datos como para que la opinión previa no influyera en tu resultado.

– *Hylas.* Y debo nuevamente preguntarte qué gano entonces con considerar esas opiniones previas, si lo que estoy intentando es precisamente librarme de ellas.

---

26. Esto es lo que modernamente se llama "método de máxima verosimilitud".

27. Filonús está aplicando de nuevo el teorema de Bayes. La probabilidad *a posteriori* de un valor es el producto de la probabilidad *a prori* (constante, en el ejemplo de Hylas) por la probabilidad de obtener los datos si el valor considerado fuera el verdadero.

– *Filonús.* Pues aunque pierdas toda opinión previa porque sean tus datos y no los de otros los que vayan a gobernar tus resultados, ganas mucho, Hylas, mucho más de lo que sospechas. Ahora puedes hablar de cuál es la probabilidad de que el peso de tus lechones esté entre éste y aquél valor, de qué probabilidad tienen de ser mayores que los de otra raza; puedes decir que tu hipótesis A es dos veces más probable que la B pero cinco veces menos probable que la C; en definitiva, puedes expresar la incertidumbre mediante probabilidades y no mediante oscuros mecanismos que invocan el lugar que ocupan fantasmales muestras que nunca tomaste y nunca has de tomar[28].

– *Hylas.* Entiendo las ventajas que tu método ofrece, pero me asaltan aún algunas dudas. ¿Qué ocurre si la hipótesis correcta no está entre las que estoy probando?

– *Filonús.* Hylas, no sabemos cuál es la correcta, ni podemos saberlo.

– *Hylas.* Te entiendo, pero ¿cómo es que puedo asignar una probabilidad a otras hipótesis, cuando la más altamente probable no está entre ellas?

– *Filonús.* Porque las probabilidades que estás comparando son relativas. Tú expresas tu creencia en que una sea más probable que la otra, pero tus resultados no hacen referencia a otras posibles hipótesis que no has considerado. O si quieres verlo de otra forma, que sí has considerado pero a las que has dado probabilidad nula. Veo que tu tendencia a imaginar lo que no haces; a tomar muestras imaginarias, digo, te hace pensar que hay infinitas hipótesis posibles que deberías considerar. No, Hylas; no puedes hablar más que de lo que hablas, no de lo que no; tus comparaciones se refieren siempre a las hipótesis que te ocupan en ese momento, y la probabilidad que tienen se refiere a cuánto es más probable una de ellas respecto a la otra.

– *Hylas.* Pero si yo considerara una nueva hipótesis, entonces las probabilidades de las demás se verían modificadas ¿no es así?

– *Filonús.* Así es, aunque no sé porqué esto te inquieta.

– *Hylas.* La razón es clara, Filonús, ¿cómo puede ser que una hipótesis tenga una probabilidad del 60%, y pase a tener el 40% si considero otras alternativas?

– *Filonús.* De nuevo crees, Hylas, que las probabilidades son algo externo, perteneciente a las cosas, y no, como ya habíamos convenido, la expresión de tu opinión sobre ésta o aquélla hipótesis. Cuando consideras alternativas nuevas, lógicamente tu grado de creencia sobre la verdad de una u otra hipótesis se modifica. Antes le diste probabilidad nula a esa nueva alternativa, y ahora tu opinión ha mudado, lo que afecta a las probabilidades de las otras hipótesis, puesto que entre todas deben sumar 1.

---

28. La imposibilidad de hablar de probabilidades para precisar la incertidumbre asociada al valor de un experimento, a menos que se admita la existencia de probabilidades *a priori*, es un asunto no discutido ni siquiera por la escuela frecuentista. Fisher, decidido anti-bayesiano, era perfectamente consciente del problema, y merece citarse un párrafo de una de sus obras (FISHER, 1936):

"Bayes percibió la importancia fundamental de este problema y propuso un axioma que, de ser cierto, bastaría para traer al campo de la probabilidad un gran número de inferencias inductivas; de forma que después de observar una muestra se podrían hacer inferencias acerca de la población; inferencias inciertas, por supuesto, pero con la bien conocida clase de incertidumbre que proviene de las leyes de la probabilidad".

– *Hylas.* Pero entonces yo no puedo establecer mi probabilidad absoluta, aunque sea subjetiva, de cualquier hipótesis. Cuando digo que su probabilidad es del 60% es sólo un espejismo, puesto que en realidad mis creencias se modificarían si yo tuviera información suficiente, y probablemente elegiría otra hipótesis más apropiada al estado de las cosas.

– *Filonús.* No sé qué quieres decir con esa probabilidad absoluta, ni tampoco con lo de tener suficiente información. No es así como nos comportamos cuando queremos averiguar el estado de la Naturaleza. Si tuviéramos esa información suficiente no sé a qué íbamos a hacer experimentos.

– *Hylas.* No me dejas convencido, Filonús, pero achácalo a los límites que Dios impuso a mi entendimiento o a la dificultad con la que lo nuevo se instala en nuestro espíritu. No quiero seguir hablando de este problema porque quiero que me contestes a otra dificultad que he visto en tu doctrina. ¿Qué ocurre cuando carecemos completamente de información previa? Si bien te entiendo, sea ésta mucha o poca, es imprescindible para que podamos hablar de probabilidades.

– *Filonús.* En rara situación me pones, Hylas, pues no sé yo cuándo has oído tú hablar sobre la ignorancia total en lo que a una ley de la Naturaleza se refiere. Novedoso experimento es éste. En ese caso yo te aconsejaría que dijeras que todas las alternativas previas tienen la misma probabilidad.

– *Hylas.* No me parece a mí, Filonús, que vaya a ser lo mismo decir que ignoro algo que decir que todas las alternativas tienen la misma probabilidad previa. Si tengo un saco en el que mi criado metió bolas blancas y negras, y no sé en qué proporción lo hizo, no creo yo que sacar una bola blanca o una negra tenga la misma probabilidad. Eso dependerá de en qué proporción estén realmente.

– *Filonús.* No estás hablando de cuál es su probabilidad real sino de tu opinión *a priori.*

– *Hylas.* Aún así no veo yo que mi opinión a priori deba ser la de equiparar todas las probabilidades. Este proceder se parece más a la indiferencia que a la ignorancia[29].

– *Filonús.* Pues entonces prueba varias opiniones previas, y si tienes suficientes datos, éstas no afectarán al resultado final, con lo que te liberas del problema porque sabes entonces que las opiniones *a priori* no van a influir en tus pronunciamientos.

– *Hylas.* Aún me queda una duda, Filonús. Cuando quiera averiguar varias cosas a un tiempo (por ejemplo, cuánto vale éste o aquél carácter, qué relación guardan entre sí) y esto lo extienda a muchas de las características de un gorrino, ¿cómo podré entonces expresar mi opinión previa? No tengo yo sesera suficiente como para pensar en cuánto puede pesar un lechón si la madre tuvo una pequeña o gran camada y al mismo tiempo la alimentaron de una forma especial, así o de otra manera, y examinar todas las variantes y relaciones entre caracteres, dando mis opiniones para cada caso. Mucha exigencia sería ésta para mí y temo que incurriría fácilmente en contradicciones.

29. No anda desencaminado Hylas. Aunque en la literatura se llaman a prioris no informativos a aquéllos en los que todas las alternativas tienen la misma probabilidad (*a priori* "planos" en el caso continuo), son todos ellos informativos. La ignorancia no puede expresarse mediante probabilidades. Una explicación más detallada se encuentra en mi artículo sobre la controversia bayesiana (BLASCO, 1998).

– *Filonús.* Te entiendo Hylas, y en ese caso no encuentro otro medio que recurrir de nuevo a probar varias opiniones y a esperar que los datos las conviertan en innecesarias.

– *Hylas.* ¡Pero es que en este caso me encuentro incapaz de expresar esas opiniones, Filonús!. No es que me resista a comparar varias creencias previas, es que no sé cómo expresarlas con exactitud.

– *Filonús.* Difícil problema es éste, y sólo puedo sugerirte que hagas lo que puedas partiendo de bases diferentes. Por ejemplo, puedes sostener por un lado que todos los valores tienen la misma probabilidad *a priori*, y por otro establecer las opiniones previas sobre cada carácter por separado, como si no estuvieran relacionados unos con otros. Puedes sostener que los caracteres están relacionados de alguna forma vaga y más o menos arbitraria. Si los resultados al final son los mismos, es difícil sostener que la opinión previa tenía algún valor, puesto que partiste de bases diferentes para construir cada opinión *a priori*.

– *Hylas.* De nuevo me parece que tu teoría pierde encanto, Filonús. Aunque reconozco, y así lo digo, que no es pequeña ventaja el que podamos aún movernos dentro del mundo de las probabilidades para expresar la incertidumbre.

– *Filonús.* Así lo creo yo también. Y déjame que muestre mi satisfacción por haberte persuadido de la gran recompensa que obtenemos, aún a costa de pagarla con imprecisiones en nuestra forma de proceder.

– *Hylas.* Así parece que Dios no construyó un mundo perfecto en el que pudiéramos decidir sin errar el cómo comportarnos. Pero me doy con esto satisfecho, Filonús, y retirémonos ya, que el sol está al caer y andamos todavía lejos de nuestras casas.

Tercer diálogo

– *Filonús.* Buenos días Hylas. Espero que te encuentres mejor y puedas disfrutar de esta limpia madrugada, tan inusual en las colinas de Cotswold, y de la belleza que muestra nuestra catedral[30] iluminada por los primeros rayos de ese Sol de cuya fidelidad tanto hemos dudado en estos días.

– *Hylas.* Buenos días. Filonús. Algo positivo traje de nuestras conversaciones, y es el agradecimiento a Dios por permitirnos que, efectivamente, luzca el Sol como esperamos, a pesar de nuestras dudas sobre ello. Aunque debo confesarte que aún me inquieta el que no podamos conocer la naturaleza de las cosas como algo cierto, sino sólo como probable. Es como si la realidad se viera tamizada a través de una neblina, o como si fuéramos ciegos y tuviéramos necesidad de un lazarillo para poder situarnos ante el mundo que nos rodea.

– *Filonús.* Pero debes pensar, Hylas, que esa probabilidad se acerca cada vez más a la certeza. Una vez has realizado un experimento, tu estado de creencias es nuevo y es distinto al que tenías antes de ponerte a trabajar. Tus dudas son menores y tu conocimiento más preciso. Y ese es el estado de opinión que tú u otros utilizarán para experimentar de nuevo, si sintieran la necesidad de precisar mejor su incertidumbre.

30. Se refiere a la catedral de Lincoln, magnífico ejemplo de gótico flamígero inglés, y cuya belleza trasciende a la imagen que presenta de madrugada.

– *Hylas.* Quieres decir que, con el tiempo, yo u otros más dotados de fortuna irán precisando mejor su conocimiento y cegando poco a poco el pozo de su ignorancia. Si te entiendo bien, tras cada experimento puedo precisar mi estado de creencias nuevamente, y cuando quiera acercarme más a la verdad utilizaré ese estado de creencias como *a priori* en el proceso de indagación, con lo que, si Dios lo permite, estaremos cada vez más cerca de la verdad y del conocimiento de las leyes que gobiernan la Naturaleza.

– *Filonús.* Así lo veo yo, y es así como podemos adquirir conocimiento general a partir de los hechos particulares que examinamos.

– *Hylas.* Si estás en lo cierto, habrías resuelto el problema de la inducción, que ha ocupado a tantos filósofos durante tantos siglos.

– *Filonús.* Pues ya que resolvimos el problema de la inducción, deleitémonos hoy con placeres más próximos a los sentidos y dejemos nuestras elucubraciones para momentos en los que la melancolía se apodere de nuestro espíritu. Tiempo para todo hay, y días buenos no tantos, así que te propongo que ya que has serenado a tu entendimiento, excitemos ahora lo que de él menos depende.

– *Hylas.* Razón tienes, y te la doy por eso; pero antes de que pueda gozar realmente de los placeres que me estás sugiriendo debe cumplirse la premisa de que mi entendimiento esté sereno.

– *Filonús.* ¿Y no es así ahora? ¿O es que nuevas dudas te asaltan sobre todo lo que tanto hemos discutido? ¿Vuelves a creer acaso que puedes salirte de ti para ser juez imparcial de cuanto experimentas?

– *Hylas.* No temas, Filonús, que todo eso me quedó bien claro. Lo que ahora me preocupa es la forma en la que resolvimos la inducción. Sostenemos que podemos conocer la probabilidad de un suceso, o de que una variable tome un valor, dados por un lado los datos obtenidos en el experimento y por otro la probabilidad que se deriva de nuestra opinión previa.

– *Filonús.* Así es, y que aunque esta probabilidad sea subjetiva, cuando el número de datos sea suficientemente grande, tu opinión previa debe ser irrelevante.

– *Hylas.* Pero es que hay algo que no me acaba de cuadrar en el proceso. ¿Qué ocurriría si yo hubiera obtenido mal la muestra? ¿Qué pasaría si los datos no reflejaran bien el estado de la población? ¿Qué sucedería si no se distribuyeran como yo presumo? ¿Qué ocurriría si mi opinión previa, compartida por muchos científicos, se alejara tanto de la verdad debido a profundos prejuicios, tan comunes entre los hombres, que ni siquiera con un número razonablemente elevado de datos pudiera contrarrestar la errónea opinión vertida por expertos?

– *Filonús.* Mira Hylas, que te estás dejando caer por la pendiente que conduce al escepticismo, y de ahí al solipsismo y la melancolía hay poco trecho.

– *Hylas.* ¿Debo, pues, engañarme y decir que entiendo lo que no, y que admito como razonable lo que me parece fundado con pobreza?

– *Filonús.* No, no debes hacerlo, pero tampoco debes no reconocer los límites del entendimiento humano. No existen probabilidades *a priori* solas, flotando entre las olas producidas por mares de opiniones. Toda probabilidad se basa en ciertas hipótesis, y también lo hace esta probabilidad previa. En realidad deberíamos llamarla "probabilidad *a priori* dadas una serie de hipótesis", pues sólo bajo ellas esta probabilidad es factible.

– *Hylas.* Pero entonces el resultado de nuestra inferencia no es una probabilidad "dada una evidencia experimental", sino que es una probabilidad "dada una evidencia experimental *y un conjunto de hipótesis anexo*".

– *Filonús.* Así lo entiendo yo.

– *Hylas.* ¿Y cómo dices entonces que has resuelto el problema de la inducción, si lo has dejado tan intacto como antes de empezar con tus elucubraciones? ¡Virgen está aún, si yo ando en lo cierto! ¿Cómo puedes afirmar que la Ciencia tiene ya una base firme, si todo depende de la verdad de unas hipótesis previas, y ésta es una verdad que ignoras? ¿Cómo puedes justificar que tu conocimiento progresa si su fundamento es tan frágil como el barro de los pies del ídolo? ¿O es que tienes algún medio de estimar la certeza de las hipótesis que acompañan a todo tu razonamiento?

– *Filonús.* Tranquilízate, Hylas, no te vaya a dar algún mal en los livianos, que primero es tu salud que todo lo del Mundo. Yo no sé qué tan ciertas puedan ser esas hipótesis, aunque si sospechara que la muestra no estaba bien tomada o que el operario que midió a los cerdos trabucó los datos por azar, o por malicia porque tenía alguna deuda conmigo pendiente; en ese caso, digo, no daría crédito a mis resultados, luego si se lo doy es porque está en mi convencimiento que todo se hizo de manera honrada y razonable. Y la ciencia así progresa, porque puedo elegir entre teorías rivales aunque estén condicionadas a la veracidad de esas hipótesis, puesto que al final la evidencia experimental me decantará por una de las teorías o por la otra.

– *Hylas.* Pobres medios me parecen para alcanzar tan elevadas metas. Si debe la ciencia progresar en base a la buena fe, a la ausencia del error inevitable, a la correcta interpretación de la Naturaleza, y a todo lo que tu bolsa de hipótesis contiene, no te diría yo que espero un progreso rápido y constante, sino altibajos, vueltas atrás y errores que quedarán sin resolver porque confiamos en que hicimos bien las cosas.

– *Filonús.* No es pobre resultado el tener un manera de discernir qué teoría preferimos, aunque es cierto que queda la ciencia siempre condicionada a la verdad de esas hipótesis. En realidad siempre pende de nosotros la espada de Damocles de estar bajo hipótesis erradas, pero una vez admitamos que son las más razonables que podemos encontrar, podemos poner en marcha nuestro mecanismo de decisión y precisar nuestras estimaciones de todo lo que hallamos en la Naturaleza, decir cuán probable es que este carácter tenga este o aquél valor, y preferir esta teoría a aquélla basándonos en cuánto más probable es ésta que sus alternativas.

– *Hylas.* En cuanto *crees que es* más probable, querrás decir.

– *Filonús.* Por supuesto Hylas, ya hemos convenido en que la probabilidad no sea otra cosa que mi estado de creencias.

– *Hylas.* Así es, pero tengo la impresión de que es tanta nuestra costumbre de creer que la probabilidad es algo objetivo y que está fuera de nosotros, que a base de hablar de cuán probable es esta o aquella hipótesis, acabemos por creer que hablamos en verdad de probabilidades objetivas y no de nuestro estado de opinión.

– *Filonús.* ¿Y qué sugieres? ¿que los científicos escriban sus comunicaciones a la Royal Society hablando de "mi probabilidad"? Sería enfadoso y acentuaría el ego del autor, algo que debe evitar siempre un hombre de ciencia.

– *Hylas.* Sí, pero no daría la impresión de una objetividad que le es ajena.

– *Filonús.* Y que tampoco pretende. También el insistir en la subjetividad daría la impresión de que se hace algo arbitrario, y ya convenimos en que subjetividad no implica para nada arbitrariedad, si el científico no es un loco. Y esto sin contar que las más de las veces los datos van a dominar sobre esa opinión previa subjetiva que tanto te repugna.

– *Hylas.* Admito que la ciencia puede progresar, aunque dudo de que ese progreso sea tan rápido y constante como la aplicación de tu principio parece sugerir, pero hemos dejado intocado el principio de inducción. En realidad ninguna evidencia experimental nos garantiza la proximidad a la verdad, puesto que dependemos de otras hipótesis cuya verdad ignoramos.

– *Filonús.* Pero no me negarás que hemos establecido un comportamiento razonable para que pueda ser seguido por todos aquellos hombres que se interesen en conocer las leyes de la Naturaleza.

– *Hylas.* Comportamiento que producirá sus buenos resultados porque Dios no permitirá que sea de otro modo, pero no porque intrínsecamente garantice una aproximación a la verdad.

– *Filonús.* Hylas, me traes a colación una comparación que pudiera ser afortunada. La ética sin Dios se basa en el comportamiento que más beneficia al individuo y a la comunidad. De la misma forma, un progreso científico sin Dios se basaría en que el comportamiento de un científico va a dar lugar a resultados que consideramos, bajo ciertas hipótesis, como más probables y que mejor se ajustan a cuanto conocemos. En ambos casos se determina el comportamiento del individuo sobre bases inseguras, pero experimentales.

– *Hylas.* No entiendo la comparación, Filonús, y me parece traída por los pelos. En lo de la ética porque no sé como armonizas el beneficio del individuo con el de la sociedad, que tan a menudo entran en conflicto; y en lo de la ciencia, porque no sé qué tiene que ver el comportamiento del científico con lo que te estoy preguntando: ¿cómo puedes asegurar que las evidencias experimentales te colocan más cerca de la verdad?

– *Filonús.* Hylas, si tan extremo eres te diré que no, que no puedo asegurar que evidencia experimental alguna me aproxime a la verdad; que tampoco puedo asegurar que tú estés aquí ahora y no esté dirigiéndome yo a un fantasma o esté inmerso en un sueño sin saberlo.¡Ni siquiera sé si puedo hablar apropiadamente como de mí mismo, puesto que de mí sólo tengo una evidencia experimental!

– *Hylas.* Te ruego que no te ofendas, Filonús, mis preguntas surgen de mi perplejidad y no de afán de ridiculizar tus propuestas. Yo quería seguridad y la cambié contigo por probabilidad, y ahora no me queda ni el consuelo de alcanzar un conocimiento probable. ¡Qué felices deben ser los monjes, que en nada dudan y en Dios confían para que gobierne y resuelva el curso de sus vidas! ¿Es esa fe ciega la que en definitiva guía nuestro convencimiento de que conocemos algo?

– *Filonús.* Hylas, para empezar a discutir hay que partir al menos de la base de que tu interlocutor está realmente presente. Añadir hipótesis a nuestros experimentos, como el que no fueron saboteados, las muestras se tomaron al azar o ningún sesgo inadvertido fue introducido, no me parece que sea pagar un alto precio por las conclusiones a las que vamos a llegar. Y si el resultado te permite actuar sobre la Naturaleza y sacarle provecho, si haces puentes y no caen, si tus cosechas mejoran, si curas a enfermos, todo esto indica que no puedes andar muy equivocado.

– *Hylas.* Sí, pero la práctica ordinaria no requiere del conocimiento especulativo[31]. Tú puedes creer que estás enfermo porque malos humores se adueñaron de tu sangre, y que tomando ciertas hierbas sanarás; y si tomas las adecuadas, la salud volverá a tu cuerpo aunque las causas de tu mal no fueran las que tú creías. Usar de teorías con escaso fundamento porque funcionan bien es como hacer creer a los niños que existe el hombre del saco, porque así tomarán la sopa. Un ingeniero puede que no necesite de más que de reglas que pueda guardar en la memoria, pero un científico se ve obligado a pensar de forma diferente y a preguntarse el porqué de las cosas que infiere y de las decisiones que toma.

– *Filonús.* No sé si el Mundo está hecho para ti, Hylas, pero creo que tenemos buenas razones para actuar como actuamos y para creer en cuanto creemos.

– *Hylas.* ¿Y cuál es la naturaleza de esas buenas razones?

– *Filonús.* No te podría responder ahora, Hylas. La alegría con que empecé el día se ha desvanecido, y veo que nuevas materias deben ser sometidas a nuestra reflexión.

– *Hylas.* Bajemos, pues, a la ciudad, y otro día continuaremos tratando de estas cosas. Liberemos al cuerpo de sus tensiones, ya que no pudimos liberar al espíritu. Hay quien dice que el pensador debe ser frugal, y que a los grandes genios la intensidad del pensamiento les hacía olvidarse del sustento. No estoy yo entre ellos y no te animo a que lo estés.

– *Filonús.* Bajemos pues HYLAS, y quédese la discusión para mañana.

Post Scriptum

Prácticamente desde la invención de la estadística, y de seguro desde que Laplace formalizó su teoría de las probabilidades, los métodos estadísticos que se aplicaron hasta los años 20 eran métodos bayesianos. En los años 20 aparecieron los libros sobre probabilidad de Reisenbach y Von Mises, que definían la probabilidad como un límite de una frecuencia cuando el número de repeticiones de un evento tiende a infinito, y los trabajos de Neyman y Pearson en los años 30 consolidaron lo que hoy se conoce como escuela frecuentista. Su intento fue el de suprimir toda referencia a la información a priori, y el trabajar con probabilidades "objetivas", no dependientes del investigador. En mi opinión la escuela frecuentista da una respuesta útil a gran parte de los problemas que comúnmente plantean los científicos, pero es notoriamente insuficiente para abordar problemas complejos como muchos de los que tratamos los genetistas animales.

La escuela frecuentista alcanza su máximo esplendor con el test t, que permite comparar hipótesis sin invocar el conocimiento de parámetros poblacionales. Para Fisher el artículo de "Student" (1906), en el que este test se proponía por primera vez, era la obra cumbre de la estadística. Otra técnica que goza de popularidad en la escuela frecuentista, el método de máxima verosimilitud, tiene la propiedad interesante de que cualquier parametrización de un estimador máximo verosímil es a su vez un estimador máximo verosímil, lo que no ocurre obviamente con las probabilidades. Para abordar problemas sencillos, desde la comparación de medias hasta los análisis de regresión, las

31. Hylas se refiere a un comentario recogido en el tercero de los diálogos que publicó el obispo Berkeley: "De ahí que la gente corriente persista en sus equivocaciones y, a pesar de ello, se las arregle para desenvolverse en los asuntos de la vida".

técnicas frecuentistas son técnicas que funcionan muy bien y que no creo que vayan a ser desplazadas por ninguna otra. Sin embargo tienen limitaciones importantes que no se pueden olvidar. Por ejemplo:

1) La incapacidad para considerar el error cometido al estimar los parámetros de dispersión (varianzas y covarianzas) al estimar los de posición (valores medios o individuales).

2) La incapacidad para elegir entre dos modelos a no ser que estén jerarquizados. Incluso en ese caso se suele recurrir al test de cociente de verosimilitudes, que presenta una inconsistencia notoria: cuando la muestra tiende a infinito, la probabilidad de la hipótesis alternativa dado que la nula es cierta, no tiende a cero. Requiere además ciertas asunciones de regularidad, frecuentemente de normalidad, tamaño de muestra suficiente etc. que no pueden testarse.

3) La incapacidad de transformar un problema multivariante en varios univariantes. Frecuentemente se está interesado en unos pocos parámetros y se desearía hablar de ellos y no del total. Un máximo en un REML multivariante es, por ejemplo, notoriamente difícil de estimar, y el máximo multivariante no tiene el porqué coincidir con los máximos univariantes.

4) La incapacidad de proveer un método universal que guíe al investigador para obtener una solución a su problema. Por ejemplo, si se desea estimar parámetros genéticos de curvas de lactación en poblaciones seleccionadas, se debería realizar un análisis multivariante con los parámetros no lineales de la curva y con los caracteres objeto de la selección. No está claro cómo abordar el problema de forma frecuentista, o se requiere un conjunto de asunciones y aproximaciones más que notable.

Por otra parte los métodos estadísticos tienden a usarse de modo rutinario, conduciendo a no pocas incongruencias. Por ejemplo, es frecuente dar la media y la desviación típica como parámetros descriptivos de los caracteres; típicamente se usa la cuasi-varianza para estimar la varianza, porque es un estimador insesgado, y a continuación se halla la raíz cuadrada para estimar la desviación típica. Pero resulta que la raíz cuadrada de la cuasi-varianza no es un estimador insesgado de la desviación típica, porque la propiedad del insesgamiento no se conserva con la mayor parte de las transformaciones. O bien, se hallan estimas máximo verosímiles con un conjunto reducido de datos (reducido por casilla, quiero decir, no necesariamente en total), olvidando que todas las buenas propiedades del método de máxima verosimilitud son asintóticas. Y qué decir de los test de hipótesis, en donde los diseños experimentales se hacen para un carácter y se aplican luego los tests a un número considerable de caracteres, con lo que casi nunca aparecen diferencias significativas cuando son realmente relevantes, o aparecen diferencias significativas meramente por azar, puesto que una de cada veinte veces por término medio debe aparecer significativo algo que no lo es.

El bayesianismo, al trabajar con probabilidades permite hacer cosas que están prohibidas en el frecuentismo, y para algunos problemas concretos especialmente complicados permite encontrar soluciones notoriamente mejores que las que provee el frecuentismo. Entre ellos están:

1) Los problemas que necesitan incluir los errores de la estimación de los parámetros de dispersión en los parámetros de posición. Esto es particularmente necesario, por ejemplo, con datos correlacionados (p.ej. animales emparentados) y bases de datos pequeñas o medianas.

2) En problemas multivariantes en los que el interés se centra en una variable o en cada variable por separado. El bayesianismo permite integrar todos los posibles valores de las otras variables ponderándolas por su probabilidad. Esto es importante ante la dificultad, por ejemplo, de encontrar máximos en funciones de verosimilitud multivariantes, y ante el hecho de que el máximo multivariante no coincide necesariamente con los univariantes. Otro ejemplo ocurre en la detección de QTLs. Los métodos frecuentistas (LOD score, etc.) son insatisfactorios porque aparecen por azar falsos QTLs al probar muchos marcadores, o porque sufren de sobreparametrización.

3) En problemas difíciles que no se sabe bien cómo abordar. Por ejemplo: Comparar curvas de crecimiento o de lactación (no lineales ni linealizables) en una población seleccionada (con lo que hay que usar los datos de la selección y el problema pasa a ser multivariante). Los métodos bayesianos permiten encontrar una ruta de trabajo: construir una función posterior, que luego se estima por MCMC (Monte Carlo Markov Chains), por ejemplo.

Una última palabra en torno a la inducción. El problema está sin resolver porque obviamente no tiene solución tal y como se plantea; es decir, no hay solución deductiva para el método de inducción. Esto resulta inquietante puesto que la experiencia muestra que el método "funciona" y la ciencia progresa, pero no queda claro cómo es esto posible. La perspectiva bayesiana permite entender cómo la ciencia progresa, a base de mejorar la opinión previa hasta que no son menester nuevos experimentos, y sin embargo explica también cómo somos dependientes de la opinión previa, que puede cambiar radicalmente la interpretación de resultados si se produce una revolución con cambio del paradigma científico.

Bibliografía

BAYES T., 1763. An essay towards solving a problem in the doctrine of chances. Phil. Trans. 53: 370-418.

BERGER J.O., WOLPERT R.L., 1984. The likelihood principle. *Institute of Mathematical Statistics. Lecture Notes - Monograph Series. Purdue University*

BERKELEY G., 1734/1990. Tres diálogos entre Hylas y Filonús. *Alianza.*

BLASCO A., 1998. La controversia bayesiana en Mejora Animal. *ITEA.* 94A: 5-41.

BUNGE M., 1981. Nuevos diálogos entre Hylas y Filonús. *En: Bunge M. 1981. Materialismo y Ciencia. Ariel.* pp. 213-224.

COPLESTON F., 1963/1985. A History of Philosophy. Vol V: Hobbes to Hume. *Bantam Doubleday Dell Publ. Inc.*

DE FINETTI B., 1973. Theory of Probability. *Wiley.*

FISHER R.A., 1936. Uncertain inference. *Proc. Am. Acad. Arts Sci.* 71:245-258.

HUME D., 1739/1983. Tratado de la naturaleza humana. *Orbis.*

JEFFREYS H., 1931. Scientific Inference. *Cambridge Univ. Press.*

PLATON, C., 387/1988. Fedro. *Gredos.*

PRESS S.J., 1989. Bayesian statistics. *Wiley.*

STIGLER S.M., 1983. Who discovered Bayes's theorem? *The American Statiscian* 37:290-296.

STUDENT, 1906. The probable error of a mean. *Biometrika.* 6: 1-25.

ITEA (2000), Vol. **96A N.° 3**, 220-224

DETECCIÓN DEL ORIGEN FILOGENÉTICO DE DIFERENTES POBLACIONES PORCINAS ESPAÑOLAS POR PIROSECUENCIACIÓN DE UNA REGIÓN DEL CITOCROMO B DEL ADN MITOCONDRIAL PORCINO

A. Clop*, G. Nyman, A. Sánchez*, J. L. Noguera**, L. Andersson*****

* Unitat de genètica i Millora Animal, Departament de Patologia i Producció Animals, Universitat Autònoma de Barcelona, 08193 Barcelona, España

** Àrea de Producció Animal, Centre UdL-IRTA, 25198 Lleida, España

*** Department of Aninal Breeding and Genetics, Swedish University of Agricultural Sciences, Biomedical Centre, 751 24 Uppsala, Suecia

RESUMEN

Doscientas treinta y cuatro cerdas de 5 razas distintas (2 rústicas y 3 comerciales) fueron analizadas para una región del citocromo B (*Cyt B*) del ADN mitocondrial (ADNmt). Para secuenciar esta región se utilizó la técnica de la pirosecuenciación. Esta técnica de secuenciación en tiempo real está basada en la detección de la reacción de luminiscencia que se produce al unirse un nucleótido específico a la cadena en extensión. El objetivo de este experimento era determinar el origen europeo o asiático de las 5 poblaciones mencionadas anteriormente. Se detectaron 3 haplotipos distintos, uno de origen europeo y dos de origen asiático. Se describe además, un nuevo haplotipo originado por mutación o por recombinación genética. Estos resultados concuerdan con trabajos anteriores y con documentos históricos.

SUMMARY

Two hundred thirty four sows from 5 different breeds (2 rustic and 3 commercial ones) were analyzed for a region located on cytocrome B (Cyt B) by using mitochondrial DNA (mtDNA). We used the pyrosequencing technique for sequencing this region. This method is based on real-time sequencing and the detection of the bioluminiscence produced when each nucleotide is incorporated in the growing DNA chain. The objective of this experiment was to determine the asian or european origin of the five pig breeds mentioned above. Three different haplotypes (one european and two asian) were found. Moreover, we describe a new haplotype created by mutation or genetic recombination. These results agree well with other published papers and with historical documents.

Introducción

Actualmente el estudio de polimorfismos genéticos está avanzando hacia el uso de SNPs (*Single Nucleotide Polymorphisms*). Este avance provoca la aparición de nuevas técnicas de análisis de SNPs más rápidas y eficientes como por ejemplo la pirosecuenciación.

En un trabajo reciente de GIUFFRA *et al* (2000) se han descrito 4 SNPs localizados en una región del *Cyt B* del ADNmt porcino. Dichos SNPs se han utilizado como marcadores para determinar el origen europeo o asiático de las líneas maternas de diferentes poblaciones de cerdos (*Sus scrofa*). La razón de utilizar SNPs localizados en el ADNmt se debe a su alta tasa de mutación que le permite ser usado frecuentemente como marcador filogenético para comparar poblaciones poco distanciadas genéticamente.

Objetivos

El objetivo principal de este trabajo es la determinación del origen filogenético de la línea materna de 5 poblaciones porcinas españolas pertenecientes a 5 razas distintas mediante el uso de la pirosecuenciación.

Material y métodos

Material animal

Se compone de 234 cerdas de 5 poblaciones distintas, 2 rústicas y 3 comerciales, todas ellas españolas.

- 5 cerdos de una línea de raza Ibérica con alta consanguinidad (IB).
- 5 cerdos de raza mallorquina (PN).
- 112 cerdos de raza Landrace de una misma línea genética (LD).
- 43 cerdos de raza Large White de una misma línea genética (LW).
- 69 cerdos de raza Pietrain de una misma línea genética (PT).

Pirosecuenciación (RONAGUI *et al.*, 1998)

Es una técnica de análisis del ADN basada en la secuenciación en tiempo real, sobre producto amplificado, de fragmentos internos de pequeño tamaño. La estrategia utilizada consiste en la detección de bioluminiscencia que se produce por la acción de la Luciferasa al incorporarse los deoxinucleótidos trifosfato (dNTP) a la cadena en extensión (figura 1). La incorporación de un dNTP libera un grupo pirofosfato (Ppi) (1a). La ATP sulfurilasa convierte el PPi en ATP en presencia de ADP (1b). El ATP junto con la luciferasa cataliza la conversión de luciferina a oxiluciferina y la emisión de luz. La detección de la luz emitida permite establecer el número y orden de los nucleótidos incorporados en la reacción de secuenciación.

Los nucleótidos se dispensan por separado y en el orden deseado, sabiendo, por tanto, cual se incorpora en cada momento y posibilitando el establecimiento de correlaciones entre el pico de luz con el nucleótido que se añade en cada momento.

El instrumento de secuenciación usa tecnología INK-JET para proporcionar con exactitud los volúmenes requeridos de los distintos reactivos y además dispone de una cámara CCD para detectar la luminiscencia.

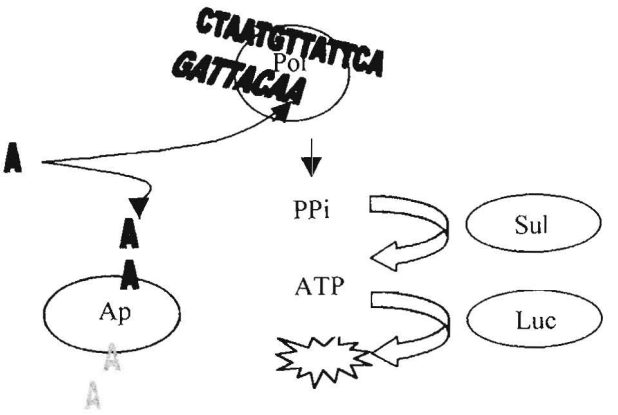


Figura 1. DNA polimerasa (Pol): *interviene en la extensión de la cadena en formación.* Sulfurilasa (Sul): *convierte el PPi en ATP.* Luciferasa (Luc): *reacciona con el ATP y produce luminiscencia.* Apiras (Ap): *Elimina el PPi, el ATP y los nucleótidos no incorporados*

Mediante la reacción de PCR, se amplificó una región de 131 pares de bases del citocromo B del ADN mitocondrial. La reacción de pirosecuenciación se realizó sobre una región de 12 pares de bases del producto amplificado que presentaba 4 SNPs.

Las reacciones de amplificación se realizaron en un volumen final de 20 µl que contenían 2 ng de DNA genómico, 3.0 mM de MgCl2, 50 mM KCl, 15 mM Tris-HCl a pH 8, 200 µM dNTPs, 0.5 unidades de AmpliTaq Gold (Perkin Elmer), 10 pmol del cebador forward (5´-GCCTACGCTATTC-TACGTTCA-3´) y 10 pmol del cebador reverse (5´-GTGGTCG AAATATTATGC - 3´). Para activar la AmpliTaq Gold se realizó una etapa inicial de activación a 95 °C durante 9 minutos, seguido de 45 ciclos con un perfil térmico de a 94 °C-45", 50 °C-45" y 72 °C-45".

Posteriormente se realizaron las reacciones de secuenciación en un volumen final de 40 µl que contenía 20 mM Tris-Acetato, 5 mM de MgCl2, 15 pmol del cebador de pirosecuenciación (5´-ATTAGGATGGAGGCT-3´) y producto amplificado purificado mediante Dynabeads. El producto resultado de la secuenciación era el descrito en la figura 2.

Producto secuenciado:

POSICIÓN: 15045 15041 15038 15036

4 SNPs: A W T A G X G C Y A Z C

C/T G/A C/T G/A

Figura 2.

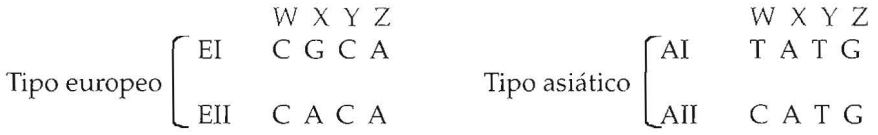


Figura 3.

La combinación de estos 4 SNPs permite caracterizar 4 haplotipos que pueden ser utilizados como marcadores filogenéticos del origen de la línea materna de los animales analizados (figura 3).

Resultados

Se presentaron 3 de los 4 haplotipos descritos en la figura 3. Además se encontró un nuevo haplotipo no descrito hasta la fecha y con un patrón mixto entre el europeo y el asiático al cual llamamos AE (ver tabla de resultados).

Discusión

De los 10 animales de razas españolas autóctonas, 9 individuos (5 IB y 4 PN) presentaron un haplotipo típicamente europeo. Estos datos concuerdan con el hecho que son animales que provienen de poblaciones muy cerradas, en las que la introgresión de animales de otras razas ha sido muy escasa. A su vez son poblaciones con una alta consanguinidad, lo que unido al bajo número de animales analizados en las dos poblaciones permite pensar en la posible existencia de sesgo en los resultados obtenidos.

Un animal PN presentó un haplotipo no descrito anteriormente (TGCA), lo que podría

Cuadro 1
Tamaños de los productos de *Overlap*-PCR de las construcciones de interacción de murinas
Table 1
Overlap-PCR sizes of the murine interaction constructs

	EI	EII	AI	AII	AE
IB	5	0	0	0	0
PN	4	0	0	0	1*
LD	94	0	18	0	0
LW	19	0	24	0	0
PT	40	0	2	27	0
TOTAL	162	0	44	27	1

(*) T G C A.

deberse a una mutación puntual en el ADNmt o a un caso de recombinación genética. De hecho, algunos trabajos recientes describen desequilibrio de ligamiento en el ADNmt de animales superiores como el hombre o el chimpancé sólo explicable por la existencia de recombinación genética (AWADALLA *et al.*, 1999).

Las líneas comerciales españolas (LW, LD y PT) presentan individuos AI que indican la presencia de genes de origen asiático (probablemente Meishan) de procedencia materna como se describe en otros trabajos y en documentos históricos (GIUFFRA *et al.*, 2000).

En el caso concreto de los animales LW se detecta una proporción mucho más elevada del tipo AI que en los animales de la población LD o PT, lo que concuerda con los hechos que fueron los británicos los que, durante los siglos XVIII y XIX, importaron un mayor número de hembras Meishan para cruzarlas con sus razas autóctonas. En la muestra de PT española se detecta la presencia elevada de animales AII, que serán secuenciados para entender la diferencia filogenética entre los dos tipos asiáticos AI y AII.

El estudio del ADNmt sólo sirve para determinar el origen de la línea materna de los animales por lo cual, en nuestro caso concreto sólo podemos indicar la introgresión de hembras asiáticas en las poblaciones europeas. No obstante, estudios recientes no descartan la posibilidad de encontrar herencia mitocondrial de origen paterno (STRAUSS, 1999). Para determinar si también se introdujeron machos asiáticos sería necesario estudiar marcadores localizados en el cromosoma Y (JOBLING, 1995).

Bibliografía

AWADALLA P., EYRE-WALKER A., MAYNARD SMITH J., 1999. Linkage desequilibrium and Recombination in Hominid Mitochondrial DNA; Science, vol. 286: 2524-2525.

GIUFFRA E., KIJAS J.M., AMARGER V., CARLBORG Ö., JEON J.T. AND ANDERSSON L., 2000. The Origin of the Domestic Pig: Independent Domestication and Subsequent Introgression. Genetics, 154: 1-7.

JOBLING M.A., TYLER-SMITH C., 1995. Fathers and sons: The Y chromosome and human evolution. Trends in Genetics, Vol. 11 N°. 11: 449-456.

RONAGUI M., UHLÉN M., NYRÉN P., 1998. Real-Time Pyrophosphate Detection for DNA Sequencing; Science vol. 281: 363-364.

WISEMAN J., 1986. A History of the British Pig. Duckworth, London.

ITEA (2000), Vol. **96A N.° 3**, 225-234

ESTIMACIÓN DE FRECUENCIAS ALÉLICAS EN *POOLS* DE ADN MEDIANTE EL PROCESAMIENTO DE SEÑALES DE ELECTROFORESIS CAPILAR

M.L. Checa*, C. Carleos, J.A. Baro***, S. Dunner*, J. Cañón***

* Laboratorio de Genética, Departamento de Producción Animal. Facultad de Veterinaria, Avda. Puerta de Hierro s/n, 28040 Madrid, España

** Departamento de Estadística e Investigación Operativa, Universidad de Oviedo, España

*** SERIDA, Principado de Asturias, España

RESUMEN

Los microsatélites constituyen en la actualidad uno de los marcadores más utilizados para la búsqueda de QTLs o loci responsables de algunas enfermedades hereditarias. Una alternativa al genotipado individual es la amplificación en una sola reacción de un grupo de animales que conforman un *pool* de ADN. En este trabajo se ha valorado la precisión en la estimación de las frecuencias alélicas de cuatro microsatélites amplificados en *pools* de ADN formados por grupos de entre 3 a 50 animales. El producto amplificado fue posteriormente analizado por electroforesis capilar en un secuenciador ABI-310. Las imágenes obtenidas se corrigieron para las bandas sombra y amplificación diferencial, artefactos que se generan durante el proceso de amplificación en la mayoría de los microsatélites, tomando como medidas informativas de las frecuencias alélicas la altura y área de pico. Como medida de la precisión del estimador de las frecuencias alélicas se utilizó la varianza del error técnico (Vt), siendo la altura de pico la variable que proporcionaba menores valores de Vt. La Vt media obtenida fue de 0,001, por lo que la frecuencia de cualquier alelo se encuentra dentro del rango $x \pm 6{,}3\%$ para un intervalo de confianza del 95%. La correlación entre las frecuencias alélicas obtenidas por genotipado individual y las estimadas a partir de los *pools* fue alta, r = 0,93 a r = 0,99. No se observó un incremento de la Vt con el aumento del número de individuos dentro del pool.

Palabras clave: Pools de ADN, Microsatélites, Frecuencias alélicas, Polimorfismo.

SUMMARY

Microsatellites are at the moment the most used marker to map quantitative trait and disease loci (QTLs). An alternative to genotype single individuals is amplifying just in one reaction a group of animals which ones make up a pool of DNA. We have quantified the precision of this technique to estimate allele frequencies for four microsatellites in *pools* from three to fifty individuals. The PCR product was analysed through capillary electrophoresis using an ABI-310 sequencer and the electropherograms were corrected for stutter bands and adjusted for preferential amplification. Both are artifactual PCR products, which are present in the majority of the microsatellites.

Peak height and area were assessed with the objective of determining allele frequencies. V_T is the technical error variance used to estimate the accuracy of prediction of allele frequency and it was lower for peak height than peak area in every used marker. The average V_T was 0.001, that is, any allele frequency x is in the range $x \pm 6{,}3\%$ for 95% confidence interval. The correlation between estimated allele frequencies from *pools* and allele frequencies based on individual genotyping was always higher than 0.93. No V_T increment was observed with the increasing number of individuals included in the *pool*.

Key words: DNA pooling, Microsatellites, Allele frequency, Polymorphism.

Introducción

El genotipado a partir de mezclas o *pools* de ADN tiene diversas aplicaciones, entre las que se encuentran la búsqueda de loci implicados en el desarrollo de enfermedades (KIROV *et al.*, 2000; DANIELS *et al.*, 1998; BARCELLOS *et al.*, 1997; SHAW *et al.*, 1998) y la búsqueda de loci asociados a caracteres cuantitativos de importancia económica o QTLs (LIPKIN *et al.*, 1998).

La amplificación de microsatélites en *pools* de ADN presenta como ventaja más evidente, frente al esfuerzo de genotipados individuales, la reducción del coste de genotipado como consecuencia de una menor necesidad de reactivos y de mano de obra.

A la hora de trabajar con microsatélites amplificados a partir de *pools* de ADN surgen problemas específicos. Por un lado será necesario una cuantificación precisa del ADN de cada uno de los individuos que van a componer el *pool*, medir la sensibilidad de la PCR para reflejar cuantitativamente las diferencias de amplificación que responden a frecuencias alélicas diferentes, medir la sensibilidad del método de cuantificación del producto de la PCR después de la electroforesis (radioactividad, quimioluminiscencia, fluorescencia...), y la repetibilidad de la técnica de elección. Además, el número de individuos presentes en el *pool* será una de las decisiones importantes si observamos que tiene una significativa influencia sobre el error de la estimación de las frecuencias alélicas.

Normalmente en los microsatélites no amplifica una única banda por alelo; a la banda principal le acompañan otras secundarias conocidas como "bandas sombra", que son el resultado de la amplificación simultánea de fragmentos entre uno y tres motivos más cortos que el alelo verdadero, consecuencia del fenómeno conocido como *slippage* (deslizamiento) de la polimerasa. Debido a estas amplificaciones accesorias, en los *pools* de ADN se produce un solapamiento de bandas principales y bandas sombra que dificulta la estimación de las frecuencias alélicas. Además, durante la PCR se producen otros dos artefactos: la adición por parte de la polimerasa de una adenina en el extremo 3' de la hebra y la amplificación diferencial que favorece a los alelos más cortos, factores éstos que también contribuyen a dificultar la estimación correcta de las frecuencias alélicas.

El objetivo del presente estudio es la formalización de las manipulaciones de los resultados de la electroforesis capilar consecuencia de la amplificación por PCR de los alelos de un microsatélite presentes en un *pool* de ADN con el fin de estimar las fre-

cuencias alélicas relativas con una precisión suficiente para ser utilizadas en trabajos de localización de QTLs.

Materiales y métodos

Extracción y cuantificación de ADN

Se extrajo el ADN de cincuenta muestras de sangre de bovino siguiendo un método de extracción fenol-cloroformo, se estimó su concentración mediante fluorimetría (TD-360 Mini-Fluorometer, Turner Designs) y se procedió a la amplificación individual de los cuatro microsatélites seleccionados con el fin de determinar el comportamiento de cada uno respecto a las bandas sombra, adición de la adenina y amplificación diferencial.

Una vez conocido el genotipo de cada uno de los animales se crearon *pools* de ADN de tres a cincuenta animales atendiendo a sus frecuencias alélicas para estudiar la relación entre el número de animales de un *pool* y el error en la estima de las frecuencias alélicas. Se hicieron dos réplicas de cada *pool* para estimar el error en la cuantificación de ADN. Una de las réplicas se amplificó dos veces para valorar la variabilidad de resultados debida a la PCR, de modo que para cada *pool* se obtuvieron tres resultados.

Microsatélites amplificados

Los marcadores estudiados fueron los siguientes dinucleótidos: BM143 (BISHOP *et al.*, 1994), RM188 (BARENDSE *et al.*, 1994), INRA003 (VAIMAN *et al.*, 1992) y TGLA126 (GEORGES M. & MASSEY J., 1992). En los tres primeros casos el cebador correspondiente al extremo 5' de la secuencia se marcó con el fluorocromo TET y el último con el fluorocromo HEX.

Electroforesis capilar de los productos PCR en ABI 310

El secuenciador automático ABI 310 usa electroforesis capilar con un polímero desnaturalizante (POP4, Perkin-Elmer Cetus) bajo condiciones de temperatura, voltaje, etc., recomendadas por el fabricante para el análisis de microsatélites. A partir del electroferograma de cada una de las amplificaciones individuales se extrajeron los datos de altura de pico y área de pico del alelo principal y de las bandas sombra accesorias, mediante el programa GENESCAN.

Correción de los artefactos de la PCR (bandas sombra, adición de la adenina en el extremo 3' y amplificación diferencial) en los pools

Cada uno de los picos observados en un *pool* es el resultado de la superposición de la banda principal y un número de picos accesorios generados por las bandas sombra de los alelos contiguos. Todos los cálculos se hicieron para altura de pico y área de pico.

La corrección de las bandas sombra se lleva a cabo de la siguiente manera:

Llamamos *altura relativa o área relativa* ($AR_{n,i}$) *de una banda sombra* al cociente $A_{n,i}/A_{n,}$ donde n es el alelo, i es el orden de la banda sombra expresado como la diferencia en número de bases con la banda principal.

Los valores de $AR_{n,i}$ fueron medidos partiendo de la información que proporcionan tanto los homocigotos como los heterocigotos en alelos con diferencia en suficiente

número de bases para que no exista ningún solapamiento de picos.

Una vez calculados los $AR_{n,i}$ de todas las bandas sombra para todos los alelos se ajustó una recta de regresión (LIPKIN *et al.*, 1998) para predecir el comportamiento de las bandas sombra de aquellos alelos no analizados. Posteriormente se corrigieron los picos de los *pools* aplicando la siguiente ecuación:

$$AC_n = A_n - \sum_{i=2}^{r} (AC_{n-i} AR_{n-i,i})$$

donde AC_n se define como la altura o área corregida para las bandas sombra y AC_{n-i} es la altura o área corregida de A_{n-i}. Por tanto para un *pool* con k bandas habrá k ecuaciones con k incógnitas AC_n.

En el caso de la banda resultante de la adición de la adenina en el extremo 3', al estar situada una base por encima del alelo principal no interfiere con la banda principal de ningún alelo en microsatélites dinucleótidos. Es posible reducir al máximo el tamaño de la banda alargando el tiempo de elongación principal de la PCR.

Queda por corregir la amplificación diferencial ya que en la PCR se favorece la amplificación de la banda principal de los alelos más cortos, al contrario que en las bandas sombra, donde los alelos más largos tienen bandas sombra más altas. Teniendo en cuenta que ambos tipos de bandas son resultado de la amplificación de los alelos que existían en el *pool*, las bandas sombra se producen en detrimento de las principales. La corrección practicada fue la siguiente:

$$AV_n = \sum_{i=-2}^{0} AC_n AR_{n,i}$$

siendo AVn el valor corregido de altura o área para bandas sombra y amplificación diferencial.

La correción expuesta es un caso particular del tratamiento de la PCR como proceso "amplificador". Mediante la teoría del procesamiento electrónico de señales (EYKHOFF, 1974; PAPOULIS, 1977) las respuestas reproducibles de un amplificador pueden ser modelizadas con precisión. Perlin (1995) considera que existe una relación lineal entre la *señal de entrada* (las frecuencias de los diferentes alelos que integran el *pool*) y la de *salida* (la lectura del secuenciador en forma de alturas o áreas de la curva). Sea x el vector de frecuencias alélicas reales (estimado arriba por los AV_n), y A la matriz formada por los elementos $AR_{n,i} / \sum_j AR_{n,j}$. Según su modelo, el vector $y = Ax$ predice las concentraciones relativas de ADN presentes en el producto de la PCR. Puesto que y es el dato observado en el experimento, se puede estimar x mediante el producto matricial $A^- y$, operación que resume los pasos detallados arriba. La matriz A^- es una inversa generalizada de A, la cual puede obtenerse por diferentes métodos (p.ej. optimización minimo-cuadrática). Los valores ajustados negativos se truncan a cero.

Finalmente, para estimar las frecuencias alélicas (FA_n) se utilizó la siguiente ecuación:

$$FA_n = \frac{AV_n}{\sum_m AV_m} \times 2N$$

siendo N el número de animales que componen el *pool*.

Como indicador de la precisión de las frecuencias alélicas se estimó la varianza de la distribución del error técnico V_T, término utilizado por otros investigadores en trabajos anteriores (LIPKIN *et al.*, 1998; DARVASI & SOLLER, 1994), como media muestral de las diferencias cuadráticas entre la frecuencia real (genotipado individual) y las frecuencias estimadas.

Otra manera de medir la precisión de las estimaciones de las frecuencias alélicas es ver en que medida nos alejamos de la recta deseable $x = y$ mediante la regresión de las frecuencias estimadas con los *pools* *(y)* sobre las frecuencias reales *(x)*.

Resultados y discusión

Intensidad relativa de las bandas sombra ($AR_{n,i}$)

En la figura 1 se muestran las observaciones y la recta de regresión de las intensidades relativas de cada una de las bandas sombra, considerando bien el valor de la altura del pico, bien el área que encierra, sobre el tamaño del alelo. En todos los casos los valores de altura proporcionan un mejor ajuste que los valores de las áreas.

Como puede observarse, las rectas de regresión para los cuatro microsatélites, tanto cuando se utiliza la altura como el área, tienen una pendiente mayor la correspondiente a $i = -2$. En el trabajo de LIPKIN *et al.*, (1998), se utilizaba como variable regresora el número de repeticiones de los alelos en lugar del tamaño final del mismo.

Estimación de las frecuencias alélicas en los pools de ADN

La varianza debido a repeticiones no resultó ser diferente de 0, tanto entre réplicas de PCR ($6{,}88.10^{-6}$) como entre réplicas de preparación del *pool* ($7{,}18.10^{-5}$), por lo que siempre se utilizó como valor estimado final de las frecuencias alélicas a partir de los *pools* la media de todas las repeticiones.

En el cuadro 1 aparecen los valores de V_T para los cuatro microsatélites cuando se utilizan los valores de altura o los de área. Los valores de V_T más bajos en todos los microsatélites (BM143, 0,0018; RM188, 0,0008; INRA003, 0,0009; TGLA126, 0,0054) se obtienen cuando se utilizan las medidas de altura de pico, lo que representa el mismo resultado, como era previsible, a partir del mejor ajuste de la regresión de las bandas sombra para altura. Resulta destacable el valor más alto de V_T obtenido para el microsatélite TGLA126, ya que es el marcador con menor número de alelos, tan sólo 3 frente a los 6-10 encontrados en los otros microsatélites. Este hecho se debe en parte a que el V_T promedia las diferencias entre las frecuencias alélicas estimadas y las veraderas dividiendo por el número de alelos. Un microsatélite tan escasamente polimórfico sería

Cuadro 1
Valores de V_T para los cuatro microsatélites cuando se utilizan la altura o el área de los picos
Table 1
V_T values for four microsatellites when high or area of the peaks are used

Microsatélites	Vt (altura)	Vt (área)
BM143	0,0018	0,0032
RM188	0,0008	0,0026
INRA003	0,0009	0,0013
TGLA126	0,0054	0,0140

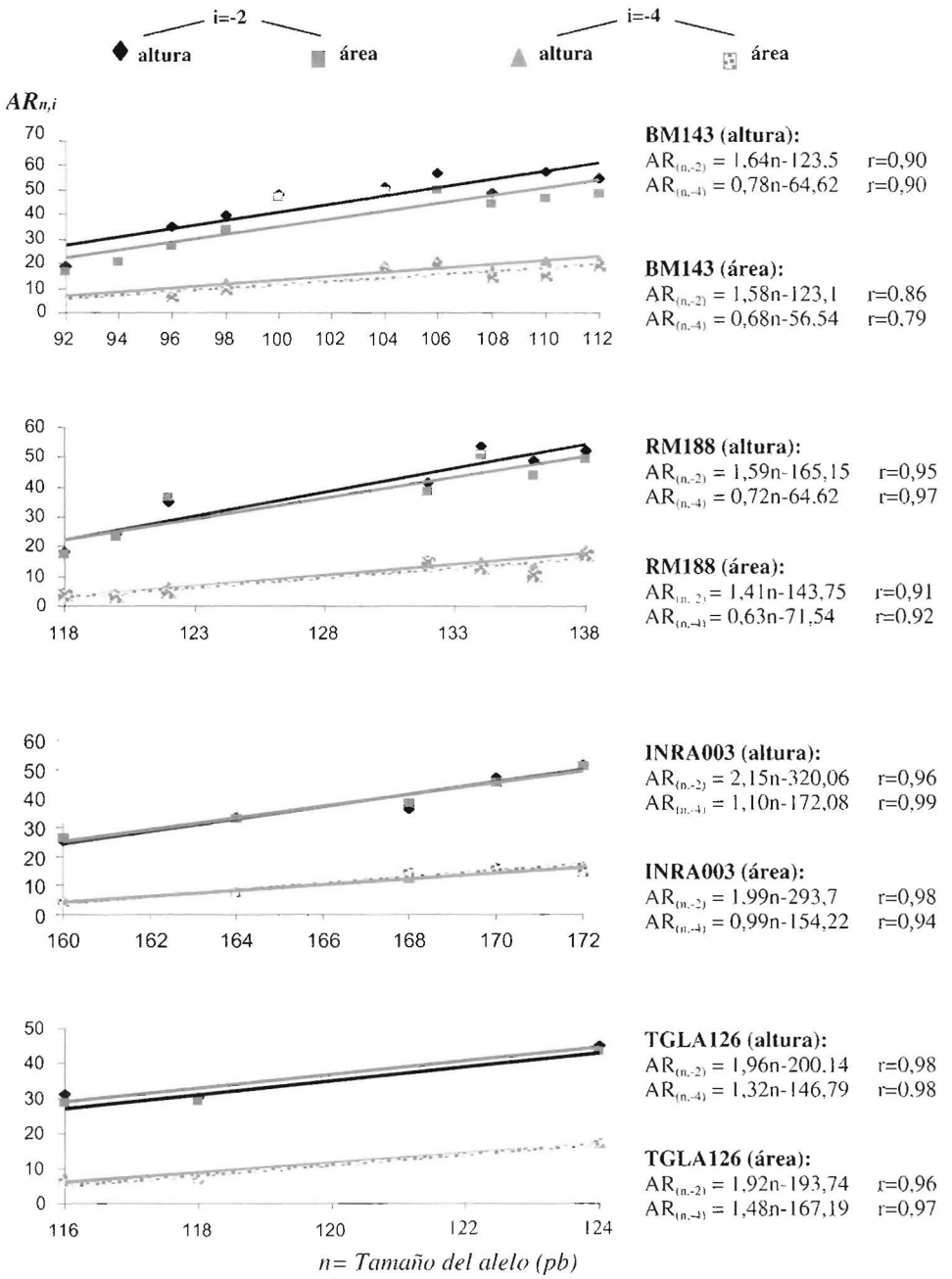


Figura 1. Regresión de las bandas sombra relativas (AR) sobre el tamaño del alelo (n). r es el coeficiente de correlación

Figure 1. Regression of the relative shadow bands (AR) on the length of allele (n). r is the correlation coefficient

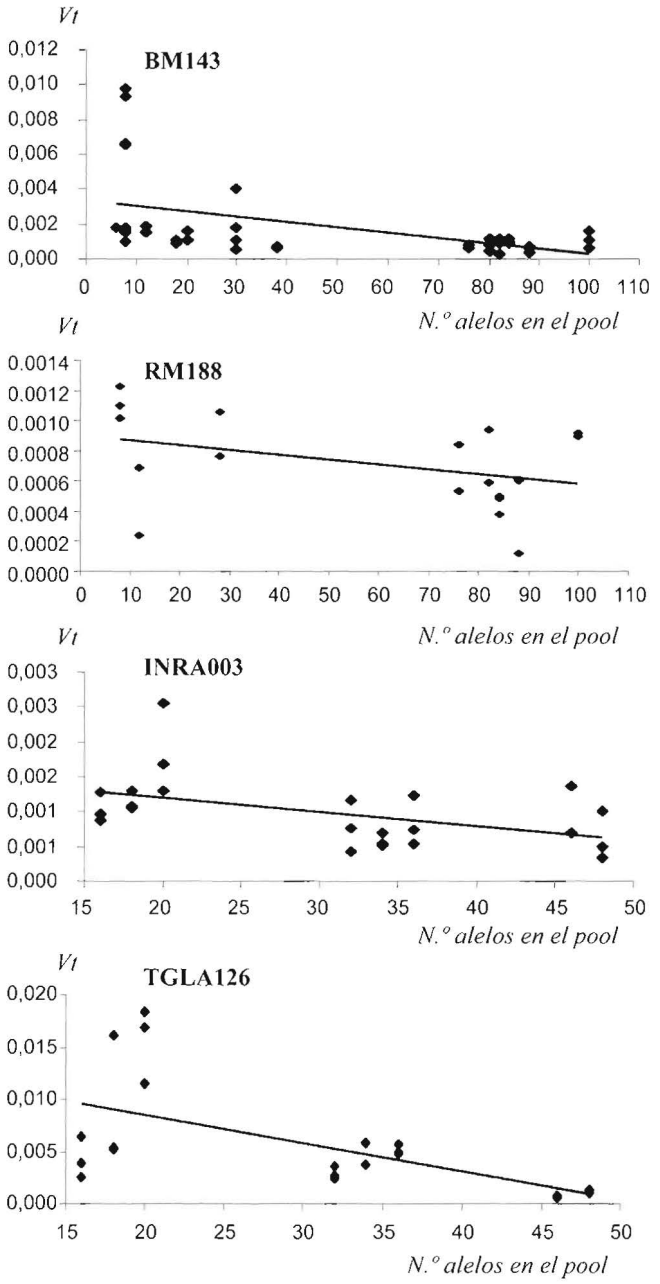


Figura 2. Evolución del error técnico (VT) a medida que aumenta el número de alelos dentro del *pool*
Figure 2. Evolution of the technical error (VT) with the increasing of the number of alleles included in the pool

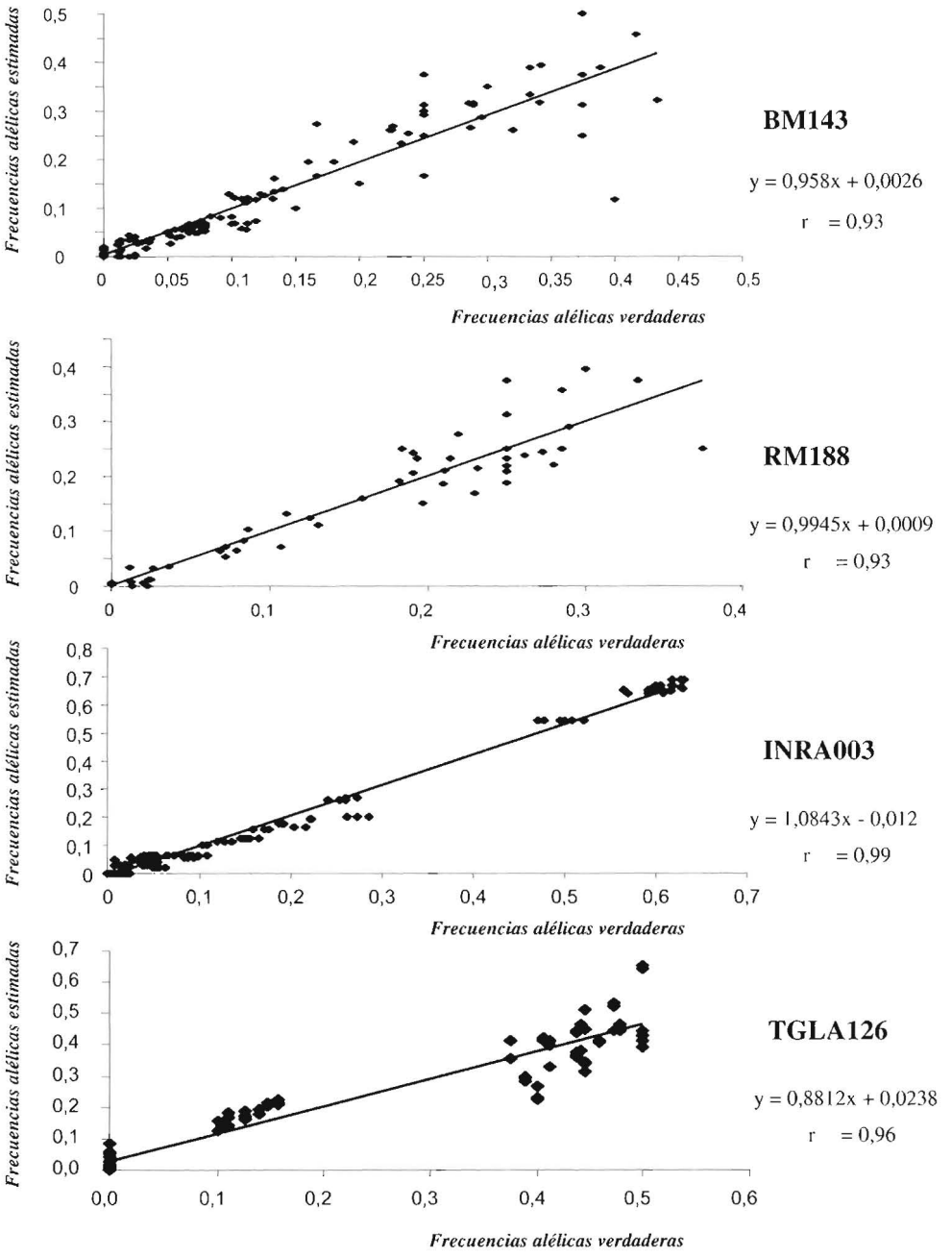


Figura 3. Regresión de las frecuencias alélicas estimadas utilizando *pools* sobre las frecuencias verdaderas en los cuatro microsatélites analizados

Figure 3. Regression of allelic frequencies estimates using pools *on the true values for the four analyzed microsatellites*

descartado *a priori*, por lo que calculamos la media de V_T con los otros tres microsatélites. Teniendo en cuenta un intervalo de confianza del 95% la frecuencia de cualquier alelo se encontraría dentro de un rango de $x \pm 6{,}3\%$. Basándonos en este error medio de muestreo ($V_I = 0{,}001$), en el caso de familias de 500 medios hermanos y un gen con efecto aditivo de 0,25 desviaciones típicas fenotípicas, el resultado obtenido se correspondería con una pérdida de potencia del 15% aproximadamente, lo que se traduciría en la necesidad de incrementar el tamaño de familia en un 26% (130 animales más) para obtener la misma potencia de detección de asociación entre el locus y el carácter cuantitativo.

En cuanto a la evolución del error técnico en función del número de individuos dentro del *pool* los resultados obtenidos demuestran que no parece que exista ningún incremento, al menos hasta un tamaño final de 50 animales que es el *pool* de mayor tamaño utilizado (figura 2).

En la figura 3 aparecen enfrentadas la estimación de las frecuencias alélicas por genotipado individual y la estimación desde los *pools*. La correlación fue alta, r = 0,93 para ambos microsatélites frente a valores obtenidos por BREEN *et al.*, (1999) entre r = 0,96 y r = 0,79. La nube de puntos alrededor de la recta se hace más dispersa a medida que las frecuencias alélicas crecen, lo que indica que la precisión en la estima es mejor para alelos cuya frecuencia de aparición dentro del *pool* es baja.

Conclusiones

Los resultados presentados proporcionan unas excelentes expectativas para justificar la utilización de *pools* de ADN en estudios de búsqueda de genes asociados a caracteres productivos o enfermedades, especialmente en las etapas iniciales de búsqueda cuando se pretende localizar regiones candidatas a lo largo de todo el genoma de una especie mediante genotipado selectivo.

Agradecimientos

Este trabajo ha sido financiado por la CICYT y Fondos FEDER a través de los proyectos: 1FD97-0042 y 2FD97-1191.

Bibliografía

BISHOP M.D., KAPPES S.M., KEELE J.W., STONE R.T., SUNDEN S.L.F., RA HAWKINS G.A., SOLINAS-TOLDO S., FRIES R., GROSZ M.D., YOO J., BEATTIE C.W., 1994. A Genetic Linkage Map for Cattle. Genetics 136:619-639.

BARCELLOS L.F., KLITZ W., LEIGH FIELD L., TOBIAS R., BOWCOCK A.M., WILSON R., NELSON M.P., NAGATOMI J., THOMSON G., 1997. Association mapping of disease loci, by use os a pooled DNA genomic screen. Am. J. Hum. Genet.61:734-747.

BARENDSE W. *et al.*, 1994. A genetic linkage map of the bovine genome. Nat. Genet. 6: 227-235.

DANIELS J., HOLMANS P., WILLIAMS N., TURIC D., MCGUFFIN P., PLOMIN R., OWEN M.J., 1998. A simple method for analyzing microsatellites allele image patterns generated from DNA *pools* and its application to allelic association studies. Am. J. Hum. Genet. 62:1189-1197.

Darvasi A., SOLLER M., 1994. Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics.138:1365-1373.

EYKHOFF P., 1974. System identification: parameter and state estimation. John Wiley & Sons. Chichester.

GEORGES M., MASSEY, J., 1992. Polymorphic DNA markers in Bovidae. Patent application WO 92/13102.

KIROV, G., WILLIAMS N., SHAM P., CRADDOCK N., OWEN M.J. 2000. Pooled genotyping of microsate-

llite markers in parent-pffspring trios. Genome Research. 10:105-115.

LIPKIN E., MOSIG M.O., DARVASI A., EZRA E., SALOM A., FRIDMAN A., SOLLER M., 1998. Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers: analysis of milk protein percentage. Genetics. 149: 1557-1567.

PAPOULIS A., 1977. Signal Analisis. McGraw-Hill. New York.

PERLIN M.W., LANCIA G., NG S.K., 1995. Towards fully automated genotyping: genotyping microsatellite markers by deconvolution. Am. J. Hum. Genet. 57:1199-1210.

SHAW S.H., CARRASQUILLO M.M., KASHUK C., PUFFENBERGER E.G., CHAKRAVARTY A., 1998. Allele frequency distributions in pooled DNA samples: applications to mapping complex disease genes. Genome Research. 8:111-123.

VAIMAN D., OSTA D., MERCIER D., GROHS C., LEVEZIET. H., 1992. Characterization of five new bovine dinucleotide repeats. Anim. Genet. 23: 537.

ITEA (2000), Vol. 96A N.° 3, 235-253

ANÁLISIS DE SUPERVIVENCIA, UNA HERRAMIENTA ESTADÍSTICA PARA DATOS DE LONGEVIDAD

V. Ducrocq*, E.A. Gómez**

* Station de Génétique Quantitative et Appliquée Institut National de la Recherche Agronomique, 78352- Jouy-en-Josas, Francia

** Departamento Producción Animal y Ciencia de los Alimentos, Universidad Cardenal Herrera CEU, 46113 Moncada, Valencia, España

RESUMEN

Los registros de longevidad tienen dos características particulares por las que requieren herramientas estadísticas especiales. Primera, parte de los datos suelen estar censurados: sólo se conoce un límite inferior de la verdadera longevidad, e.g. cuando los animales están aún vivos. Bajo ciertas asunciones es posible definir la correcta contribución a la verosimilitud de tales registros. Por consiguiente, la metodología de verosimilitud puede ser aplicada de manera estándar. En segundo lugar, las variables explicativas son a menudo dependientes del tiempo. El concepto de función de riesgo, definida como el límite de la tasa de mortalidad en el tiempo t condicionada a la supervivencia en el tiempo t, permite la modelización de las relaciones entre la longevidad y estas covariables. En modelos de riesgos proporcionales, se describe la función de riesgo de un animal como el producto de una función de referencia y una función exponencial de las covariables. Esta formulación permite una interpretación intuitiva del efecto de cada covariable en términos de riesgo relativo de muerte. Se han propuesto extensiones del método a modelos mixtos, incluyendo efectos genéticos, así como métodos para la estimación de los componentes de varianza correspondientes. Tras una extensa presentación de las bases teóricas, se presentan dos ejemplos para ilustrar las características y los usos potenciales del análisis de supervivencia: la evaluación genética de toros basada en la longevidad de las hijas y un análisis del efecto de aparición de enfermedades sobre el desecho.

Palabra clave: Análisis de supervivencia, Longevidad, Modelo de riesgos profesionales, Modelo de Cox.

SUMMARY

Longevity records have two unique features requiring special statistical tools: first, part of the data is often censored: only a lower bound of the true longevity is known, e.g., when the animals are still alive. Under certain assumptions, it is possible to define the correct likelihood contribution of such records. Then, standard likelihood methodology can be applied. Second, explanatory variables are often time-dependent (their value changes with time). The concept of hazard function, defined as the limiting rate of death at time t, conditional upon survival to time t, allows a natural modeling of the relationship between longevity and these covariates. In proportional hazards models, the hazard function of an animal is described as the product of a baseline function and

an exponential function of the covariates. This formulation allows an intuitive interpretation of the effect of each covariate in terms of relative risk of dying. Extensions to mixed models, e.g., with inclusion of genetic effects have been proposed, as well as methods for the estimation of the corresponding variance components. After a long presentation of the background, two examples are presented to illustrate the features and potential uses of survival analysis: a genetic evaluation of bulls based on the longevity of their daughters and an analysis of the effect of disease occurrence on culling.

Key words: Survival analysis, Longevity, Proportional hazard models, Cox model.

Introducción

El objetivo de los análisis de supervivencia es analizar las medidas positivas que describen la "amplitud" del intervalo entre un punto origen y un punto final. A menudo, el punto final (normalmente llamado "fallo") corresponde a la muerte o al desecho y la longitud desde el origen hasta el fin se mide en días, meses, años. Pero el punto final puede también corresponder a la aparición de cualquier tipo de suceso, v.gr., recuperación después de una enfermedad o éxito tras una inseminación, etc.

La motivación para un tratamiento estadístico especial de los datos de supervivencia se ilustra mucho mejor con un ejemplo. En la figura 1 se representa la tasa de gallinas ponedoras todavía vivas a t días del nacimiento en cuatro grupos con unas 6.000 aves cada uno.

La particularidad de este ejemplo es que el punto final (muerte) no fue observado en

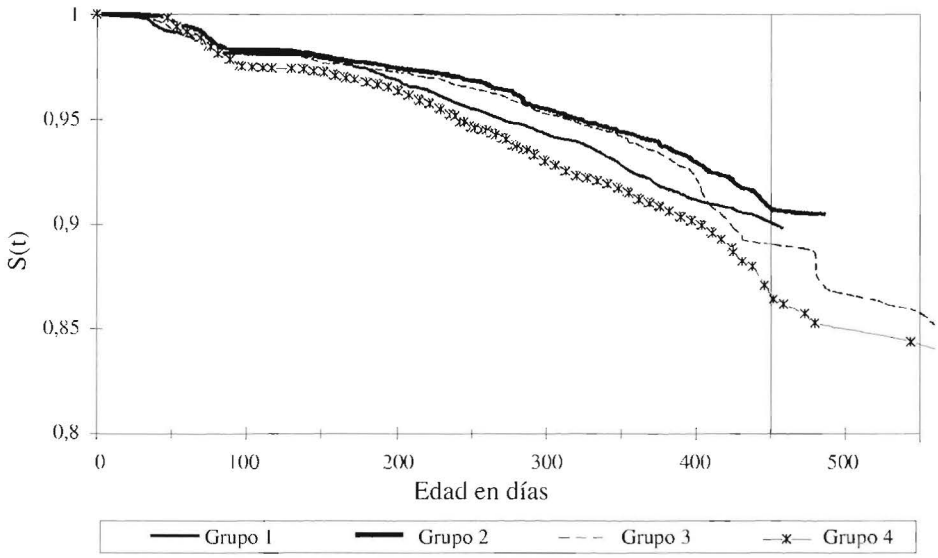


Figura 1. Curvas de supervivencia de 4 grupos de gallinas ponedoras

el 85-90% de las gallinas: están todavía vivas al final del período en estudio. Sus registros (en los que la fecha de baja se reemplaza por la fecha final del estudio) todavía proporcionan una inestimable información: da una idea del límite inferior de su "verdadera longevidad". Este fenómeno se denomina censura y el análisis de supervivencia puede ser descrito grosso modo como el arte de combinar observaciones censuradas y no censuradas en un único análisis (KALBFLEISCH y PRENTICE, 1980).

Para subrayar las ventajas de un tratamiento estadístico apropiado, consideremos la alternativa más común para analizar dichos datos: con objeto de evitar la censura, podemos colocarnos en un valor particular t de la escala temporal transformando cada registro de datos en un carácter 0-1, según la gallina esté o no viva en el tiempo t. Además, la elección de t no es trivial: con t = 100 días las diferencias entre grupos en porcentaje de supervivientes son pequeñas y poco claras, con t = 500 no disponemos de datos en dos de los grupos a los que no se les permitió alcanzar este punto final, con t = 450 la ordenación de los grupos es clara, pero no es la misma que con t = 400. Los grupos 1 y 3 tenían supervivencias similares hasta t = 400 pero por alguna razón la tasa de supervivencia fue mayor en el grupo 1 desde entonces. Los análisis de supervivencia utilizan de la manera más eficiente toda la información disponible a lo largo de las curvas de supervivencia y no se restringen a observaciones puntuales arbitrariamente definidas.

Se presentan a continuación los conceptos básicos sobre los cuales se construye el análisis de supervivencia, y cómo estos conceptos pueden adaptarse y extenderse a situaciones con modelos cada vez más complejos. Los principales objetivos serán: 1) describir cómo se combina realmente la información de los registros censurados y no censurados y 2) de donde provienen las mejoras sustanciales frente a métodos alternativos de análisis de longevidad.

Métodos

Los análisis de supervivencia usan los mismos conceptos fundamentales que otros campos de la estadística. En particular, se asientan sobre el principio de verosimilitud para la estimación de parámetros que describen los datos de modo paramétrico o no paramétrico. Se utilizarán técnicas de máxima verosimilitud siguiendo los mismos procedimientos y conduciendo a las mismas propiedades que en los problemas de la estadística clásica.

Función de riesgo y construcción de la verosimilitud

Una dificultad conceptual es la definición de la contribución a la función de verosimilitud de una observación censurada. Con objeto de derivar esta contribución, es útil la definición de las funciones específicas, que están directamente relacionadas con las funciones de distribución más usuales, como la función de distribución acumulada

$$F(t) = \text{Prob}\,[\,T < t] \qquad [1]$$

en que T representa el tiempo de fallo real, y la función de densidad f(t):

$$f(t) = \lim_{\partial t \to 0} \frac{\text{Prob}\left[t \le T < t + \partial t\right]}{\partial t} = \frac{d\,F(t)}{dt} \qquad [2]$$

Las funciones específicas más importantes utilizadas en análisis de supervivencia son las funciones de supervivencia S(t)

$$S(t) = \text{Prob}\,[\,T \ge t] = 1 - F(t) \qquad [3]$$

y la función de riesgo $\lambda(t)$ que indica la tasa instantánea de fallos (muerte) en el tiempo t condicionada a la supervivencia hasta t

$$\lambda(t) = \lim_{\partial t \to 0} \frac{\text{Prob}\left[t \le T < t + \partial t \mid T \ge t\right]}{\partial t} \quad [4]$$

$\lambda(t)$ mide el riesgo de muerte en el tiempo t, entre los animales vivos en ese instante.

Todas estas funciones están relacionadas. En particular:

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d \log S(t)}{dt}$$

$$\text{y} \quad S(t) = \exp\left\{-\int_0^t \lambda(u)\, du\right\} \quad [5]$$

Es esencial recordar las relaciones expresadas en [5]: la modelización de la función de riesgo no es fundamentalmente diferente de la de una función de densidad, siendo posible pasar de una a otra. Además, la primera es más conveniente: por ejemplo, si la función de riesgo permanece constante en el tiempo ($\lambda(t) = \lambda$ = constante), la segunda expresión de [5] indica que $S(t) = \exp\{-\lambda t\}$, que es la función de supervivencia de una distribución exponencial. Contrastar la asunción de que $\lambda(t)$ es constante en el tiempo parece más fácil que contrastar si la función de densidad es de la forma $f(t) = \lambda \exp\{\lambda t\}$.

Otra importante definición necesaria es la de censura aleatoria: podemos imaginar que para cada animal i hay un tiempo de fallo T_i y un tiempo de censura C_i aunque sólo uno de los dos es observable: T_i para una observación no censurada (en este caso sabemos que el tiempo de censura hubiera ocurrido después de $t = T_i$), C_i para una observación censurada (en este caso, sabemos que la muerte hubiera ocurrido después de $t = C_i$). La variable aleatoria a analizar es $Y_i = \min(T_i, C_i)$, i.e., el momento en que de los dos sucesos (fallo o censura) uno ocurre.

En el caso de censura aleatoria se asume que T_i y C_i son independientes. Una situación típica en que no es válida esta asunción es cuando los animales son excluidos del experimento de supervivencia cuando presentan señales de que morirán pronto. Si asumimos censura aleatoria, subestimaremos el riesgo en estos animales.

En el caso de censura aleatoria (¡y sólo entonces!), la distribución de los tiempos de censura C_i no proporciona ninguna información sobre la distribución de T_i y puede demostrarse que la contribución a la verosimilitud será:

- Para una observación no censurada $f(y_i)$ = valor de la función de densidad en el instante del fallo y [6]
- Para una observación censurada $S(y_i)$ = valor de la función de supervivencia en el instante de censura.

Las contribuciones condicionalmente independientes de todos los registros se multiplican para obtener la función de verosimilitud completa. Si se elige una forma paramétrica para f(.) o de manera equivalente para $\lambda(.)$, es posible la estima de los parámetros relevantes de la función maximizando la función de verosimilitud o su logaritmo. En muestras de gran tamaño, las propiedades de las estimas máximo verosímiles se mantienen.

Las funciones de distribución clásicas utilizadas para analizar poblaciones homogéneas se detallan en KALBFLEISCH y PRENTICE (1980), COX y OAKES (1984), KLEIN y MOESCHBERGER (1997). Se incluyen la función exponencial (con función de riesgo constante), la distribución gamma, la distribución de Weibull (una generalización de la distribución exponencial con dos parámetros: $S(t) = \exp\{-(\lambda t)^p\}$) y la distribución log-normal.

Estimación no paramétrica

Un método típico para una elección adecuada de la forma de la función de densidad f(.) de una variable aleatoria es estimar su función de distribución acumulada de manera empírica (o de manera equivalente su función de supervivencia empírica), utilizando [3] y comparando su forma con las distribuciones paramétricas conocidas. Estas distribuciones paramétricas deben ser adaptadas para considerar la censura. Sea $T_{[k]}$ los tiempos de fallos ordenados en el tiempo. Para cada $T_{[k]}$ podríamos contar cuántos animales estaban "en riesgo" (podían haber muerto en $T_{[k]}$) y cuántos murieron realmente. Sean estos números n_k y d_k. La relación entre n_k y n_{k+1} depende de la censura: por ejemplo, si en un tiempo de fallo $T_{[k]}$, n_k=100 animales en riesgo, 5 mueren y 10 se censuran entre $T_{[k]}$ y $T_{[k+1]}$, el número de animales en riesgo en $T_{[k+1]}$ será n_{k+1}=85. La función de supervivencia empírica puede calcularse utilizando la fórmula de KAPLAN-MEIER (KAPLAN y MEIER, 1958):

$$\hat{S}_{KM}(t) = \prod_{k \mid T_{[k]} < t} \left(\frac{n_k - d_k}{n_k} \right) \qquad [7]$$

La estima de la función de supervivencia de Kaplan-Meier juega un papel clave cuando queremos contrastar si la densidad de una variable tiempo de fallo pertenece a una familia en particular de las funciones paramétricas. Por ejemplo, si una gráfica del log $\hat{S}_{KM}(t)$ frente a t presenta una línea recta con pendiente λ desde el origen, podremos concluir que puede utilizarse una distribución exponencial con parámetro λ para ajustar los datos. De manera similar, un test gráfico para la adecuación a una función de distribución de Weibull consiste en contrastar una gráfica

$$\log \left[-\log \hat{S}_{KM}(t) \right] \text{ frente a } \log t$$

si da una línea recta (la pendiente y el punto de corte en ordenadas coincide con funciones de los parámetros λ y ρ de Weibull). Estos tests se han realizado para los datos de la Figura 1. En el caso estudiado, no se recomendaba la función de Weibull si se tomaba el nacimiento como punto de origen, pero sí que era asumible una función de Weibull cuando el punto de origen era el momento en que la gallina era alojada (figura 2 y DUCROCQ *et al.*, 2000).

Modelos de regresión

Sea $x = ((x_1 .. x_n)'$ un vector de variables explicativas de las que podría depender el tiempo de fallo. Las x_j pueden ser covariables continuas o discretas. Se han propuesto diferentes vías para asociar la función de riesgo $\lambda(t)$ y x. La más popular se basa en el concepto de modelo de riesgos proporcionales (PHM; COX, 1972), en los que:

$$\lambda(t ; x) = \lambda_0(t) \exp\{x'\beta\} \qquad [8]$$

siendo β un vector de coeficientes de regresión, $\lambda_0(t)$ es la función de riesgo base y representa el proceso de envejecimiento de la población total (de manera conceptual, actúa como una media). La segunda parte, $\exp\{x'\beta\}$ representa un término dependiente de las características propias de los animales con covariables x. Es siempre positiva y actúa de manera multiplicativa sobre la función de riesgo base.

Dada la relación entre $\lambda(t)$ y f(t) obtenida en [5], la expresión [8] es otra manera de definir el modelo como si se tratara de un análisis de regresión regular (mediante la caracterización de la distribución f(.) del vector residual $e = y - E[y] = y - x'\beta$). El nombre de modelos de riesgo proporcionales proviene de la siguiente propiedad: sean $\lambda_A(t)$ y $\lambda_B(t)$ las funciones de riesgo de dos

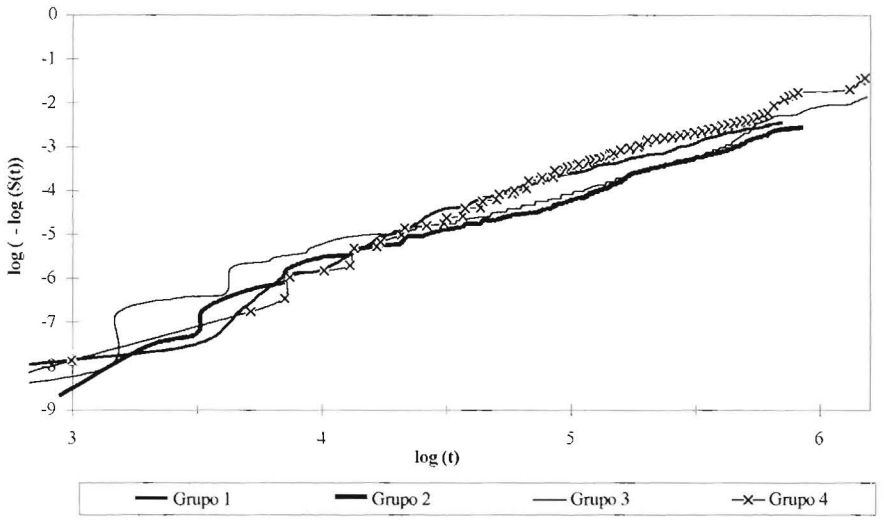


Figura 2. Test gráfico para la asunción de la función Weibull
(los mismos grupos de ponedores que en figura 1, t = días desde que se alojan,
S(t)= estima de Kaplan-Meier de las curvas de supervivencia para cada grupo)

animales con vectores asociados de covariables x_A y x_B respectivamente. El cociente de estos dos riesgos

$$\frac{\lambda_A(t)}{\lambda_B(t)} = \frac{\lambda(t\,;\,x_A)}{\lambda(t\,;\,x_B)} = \frac{\lambda_0(t)\exp\{x_A'\beta\}}{\lambda_0(t)\exp\{x_B'\beta\}} = \quad [9]$$

$$= \exp\{(x_A - x_B)'\beta\} = \text{constante}$$

no depende del tiempo. Las funciones de riesgo de estos animales son proporcionales. Si la constante es igual a 2, en cualquier momento, el animal A tiene el doble de probabilidad de fallo que el animal B. Esta simple interpretación intuitiva de los coeficientes de regresión en los modelos de riesgos proporcionales explica su gran popularidad.

Para la función de riesgo base $\lambda_0(.)$ en [8] podemos elegir cualquier forma paramétrica (exponencial, Weibull, lognormal, etc.). Además, es relativamente fácil derivar, computar y maximizar la función de verosimilitud completa, usando las contribuciones definidas en [6] para registros censurados y no censurados. Esto proporciona estimas máximo verosímiles de los parámetros de la densidad base y de β. También es posible dejar a $\lambda_0(.)$ totalmente arbitraria. La expresión [8] define entonces un modelo de regresión semiparamétrico conocido como modelo de Cox (Cox, 1972). La característica más atractiva de este modelo es que permite la estimación de los coeficientes de regresión en β sin hacer asunciones sobre la forma de $\lambda_0(.)$. El procedimiento desarrollado por Cox se basa en la definición de la función de verosimilitud parcial que es la parte de la función de verosimilitud completa que no depende de $\lambda_0(t)$. La verosimilitud parcial también admite otras justificaciones formales. En particular, puede ser obtenida como la verosimilitud marginal de los rangos de los tiempos de fallo y contiene toda la información sobre el orden en que los animales mueren. Una vez se ha estimado β maximizando la verosimilitud parcial, se puede computar una estima de la función de

riesgo de referencia $\lambda_0(.)$ o de la función de supervivencia base $S_0(.)$ (asumiendo que $\beta = \hat{\beta}$ y usando un método similar al que conduce a la estima de Kaplan-Meier $S_{KM}(.)$). Permite así el uso de contrastes gráficos (como se presentó previamente) para contrastar la validez de una forma paramétrica particular para la función base.

En algunos casos, podría ser inadecuado asumir una función de riesgo base común para todos los animales, en los modelos de riesgo proporcionales. En el ejemplo de las gallinas ponedoras, si aparece una enfermedad o se produce un súbito cambio de temperatura en un grupo, la asunción de una función base común para todos los grupos no es válida. En esta situación, es posible estratificar los datos en S estratos y asumir una función base diferente $\lambda_{0,s}(.)$ para cada estrato s (s=1,...S)

$$\lambda(t\,;\,x,\,s) = \lambda_{0,s}(t)\exp\{x'\beta\} \qquad [10]$$

De nuevo, la función base para cada estrato puede tener una forma paramétrica conocida o ser arbitraria. En la construcción de la verosimilitud (posiblemente parcial), las contribuciones de cada estrato se combinan. El uso de la estratificación conduce a un contraste gráfico de las asunciones de riesgos proporcionales: si la asunción es verificada, las $\lambda_{0,s}(.)$ son proporcionales. En las gallinas ponedoras, un modelo estratificado indica que las asunciones de riesgos proporcionales se mantienen entre grupos (DUCROCQ *et al.*, 2000).

Covariables dependientes del tiempo

Se ha venido asumiendo que las variables x no cambian con el tiempo. Es por ejemplo el caso para covariables independientes como el sexo (en muchas situaciones), raza, grupo o rebaño, peso inicial, edad al primer parto. Pero otras covariables de interés varían con el tiempo. Por ejemplo, una variable como "año" o "estación" influyen sobre la variable tiempo de fallo porque recogen todos los cambios climáticos, económicos y epidemiológicos ocurridos desde un período hasta el siguiente. De forma similar, si queremos estudiar el efecto de una enfermedad sobre el tiempo de muerte, será casi siempre incorrecto asumir que el animal está enfermo desde el día 1 hasta la muerte. El modelo de riesgos proporcionales puede extenderse a situaciones en que las covariables se modifican en el tiempo (KALBFLEISCH y PRENTICE, 1980; COX y OAKES, 1984; KLEIN y MOESCHBERGER, 1997).

$$\lambda(t\,;\,x(t)) = \lambda_0(t)\exp\{x(t)'\beta\} \qquad [11]$$

donde x(t) es el vector de covariables siendo algunas de ellas función del tiempo. Para facilidad computacional, nos autorrestringiremos al caso particular de covariables dependientes del tiempo que cambien a nuevos valores en momentos específicos y que permanezcan constantes por tramos. Por ejemplo, un variable x_j podría representar la variable que indica si un animal está enfermo (GRÖHN *et al.*, 1997):

- $x_j(t) = 0$ si el animal está sano en el instante t, v.gr., entre el punto de origen de los datos de aparición de la enfermedad y también desde la fecha de recuperación hasta la nueva aparición de la enfermedad o la muerte o la censura.

- $x_j(t) = 1$ si el animal está enfermo en el instante t.

La expresión [11] implica que la asunción de riesgo proporcional [9] se mantiene dentro de cada intervalo en el que los $x_j(t)$ permanecen constantes; pero el cociente entre riesgos puede cambiar de un intervalo al siguiente. Esta extensión es muy importante. En mi opinión, representa el mayor

avance en el arte de modelizar datos de supervivencia porque permite una descripción precisa y detallada de las variaciones a lo largo del tiempo del ambiente que conduce al fallo.

Presentaremos otro ejemplo para ilustrar esta característica. Es de todos conocido que las políticas de desecho en rebaños de vacuno lechero están cambiando continuamente: en cada rebaño dependen de consideraciones económicas (precio de la leche y de las vacas de desecho), de cambios en el tamaño del rebaño (en rebaños en expansión las vacas son eliminadas con menos celeridad) que puede depender de la cuota de referencia, disponibilidad de novillas para reemplazo, disponibilidad de forraje, etc. En los análisis de los caracteres más clásicos como producción de leche o tipo los cambios en las condiciones ambientales en un rebaño a lo largo del tiempo son tenidos en cuenta por el efecto de año-rebaño (-estación) o grupo de contemporáneas en que se registró el dato. Esto no es posible con un carácter como longitud de la vida productiva: asumir un solo grupo de contemporáneas no tiene sentido porque animales con muy diferentes longitudes de vida son contemporáneos solamente durante un corto período de tiempo. Pero si definimos año-rebaño como dependiente del tiempo, entonces parece posible asociar tantos efectos año-rebaño como sea preciso a cada registro: 2 si el animal vive 2 años, 10 si falla cuando tiene 10 años. Todo ello conduce a una modelización precisa de los cambios en la política de desecho a lo largo del tiempo, como aparece en la figura 3.

La inclusión de covariables dependientes del tiempo en modelos de riesgos proporcionales es una herramienta muy poderosa para la modelización aunque tiene dos inconvenientes:

• Hace que la interpretación sea muy difícil en algunas ocasiones (KALBFLEISH y PRENTICE, 1980); en particular cuando el cambio de la variable dependiente del tiempo está relacionado con el proceso que con-

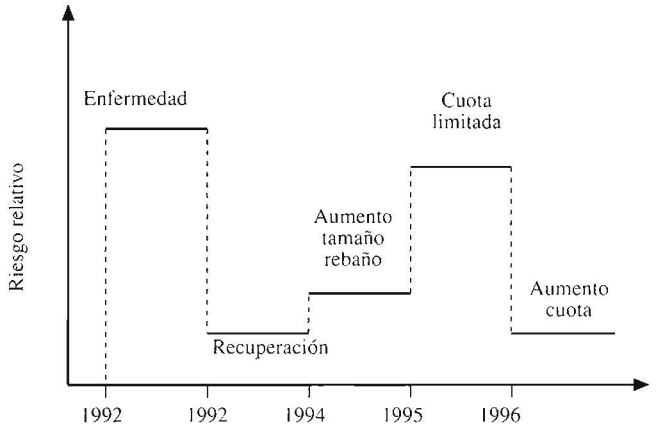


Figura 3. Ilustración sobre la modelización de los cambios en la política de desecho a lo largo del tiempo (v.gr., debido a los sucesos indicados en el gráfico) usando un efecto año-rebaño dependiente del tiempo

duce al desecho. Por ejemplo, si estudiamos el efecto de una enfermedad sobre el desecho de vacas lecheras, parece razonable corregir por producción de leche: es menos probable eliminar las mejores vacas que las peores. Si la incidencia de la enfermedad es mayor en las mejores, la no corrección por producción de leche puede conducir al siguiente resultado absurdo: la aparición de la enfermedad "protege" a la vaca frente al desecho... Por otro lado, si la consecuencia de la enfermedad es solamente una reducción brusca de la producción de leche y se elimina a las vacas enfermas por baja producción, la inclusión de la producción de leche como factor de corrección en el modelo puede conducir igualmente a la estúpida conclusión de que la enfermedad no tiene efecto sobre el desecho. Por tanto, un análisis apropiado debería incluir ambos tipos de modelos, una comparación de los resultados y un estudio de las pautas de cambio en el tiempo de la variable dependiente del tiempo. Entonces, conclusiones muy relevantes pueden ser indicadas: v.gr., el efecto de la enfermedad sobre la eliminación es importante, pero es completamente (o parcialmente, o en absoluto) debido a sus efectos sobre la producción de leche (GRÖHN *et al.*, 1998; BEAUDEAU *et al.*, 2000).

• Hace que la computación sea más dificultosa. Los programas comerciales como SAS o BMDP pueden ser extremadamente ineficientes cuando tratamos con covariables dependientes del tiempo, en particular porque son muy generales. Esto ha limitado su uso durante mucho tiempo, incluso para problemas de tamaño moderado y ha sido la principal motivación para el desarrollo de un software específico, el "Survival Kit" (DUCROCQ y SÖLKNER, 1994, 1998b) para su aplicación a gran escala en la mejora genética animal.

Modelos de fragilidad

Los modelos de riesgos proporcionales, paramétricos o no, pueden ser extendidos para la inclusión de efectos aleatorios (v.gr., genéticos) como en los modelos mixtos habituales. Los estadísticos suelen referir los modelos mixtos de supervivencia como modelos de fragilidad (VAUPEL *et al.*, 1979; CLAYTON y CUZIK, 1985).

El término de fragilidad v_m se define como una variable aleatoria no observada que cambia de forma multiplicativa el riesgo de los individuos o grupos de individuos. Cuando se define un v_m para cada individuo m:

$$\lambda(t\,;\,x,\,m) = \lambda_0(t)\, v_m \exp\{x'\beta\} \quad [12]$$

el componente de fragilidad extrae parte de la variación no observada entre individuos y por tanto permite la corrección de la posible discrepancia (sobredispersión) entre la verdadera varianza de las observaciones y la especificada por el modelo. Cuando v_m se define para un grupo de individuos, v.gr., todas las hijas de un macho m, describe las características compartidas no observables (genéticas, en este caso) que actúan sobre el riesgo de cada miembro del grupo. En todos los casos, una sencilla transformación $s_m = \log v_m$ permite la inclusión del término de fragilidad en la parte exponencial de los modelos de riesgos proporcionales. Si z representa el vector de incidencia para los efectos aleatorios $s=\{s_m\}$, el modelo mixto de supervivencia puede escribirse como:

$$\lambda(t\,;\,x,\,z) = \lambda_0(t) \exp\{x'\beta + z's\} \quad [13]$$

$\lambda_0(.)$ puede tener forma paramétrica o ser arbitraria. La expresión [13] puede extenderse para incluir funciones de riesgo base estratificadas o covariables dependientes del tiempo y/o aleatorias.

De manera tradicional, se ha supuesto una función gamma para el término de fragilidad v_m (o de manera equivalente una distribución log-gamma para el efecto aleatorio s_m) debido a su flexibilidad y conveniencia matemática. También se han propuesto otras distribuciones, v.gr., distribuciones estables positivas o distribuciones gausianas invertidas (HOUGAARD, 1986). Desgraciadamente estas distribuciones carecen del atractivo teórico de la distribución normal comúnmente utilizada en mejora animal cuando se asume un modelo poligénico infinitesimal. Sin embargo, se ha demostrado que las estimas obtenidas para los parámetros de la función gamma de v_m eran muy grandes, al menos en vacuno lechero (DUCROCQ *et al.*, 1988b). La distribución gamma tiende a una distribución log-normal a medida que los parámetros aumentan de valor (KALBFLEISH y PRENTICE, 1980) y s_m puede considerarse (al menos aproximadamente) normalmente distribuido. Por tanto, se ha sugerido una distribución normal multivariante de $s=\{s_m\}$ para tener en cuenta las relaciones genéticas entre animales (DUCROCQ, 1987).

Distintos enfoques han sido propuestos para la estima de los parámetros de la función de fragilidad: KLEIN (1992) sugería el uso del algoritmo EM, con estimación iterativa de $v=\{v_m\}$, los coeficientes de regresión β y la distribución base de riesgo acumulado con un modelo Cox, seguido por la estimación de los parámetros de la distribución de fragilidad dado $\hat{v}$. Cuando se combina un modelo Weibull con un término de fragilidad gamma, FOLLMANN y GOLDBERG (1988) mostraron que el término de fragilidad puede ser integrado algebraicamente fuera de la función de verosimilitud. La misma propiedad ha sido utilizada en un contexto bayesiano por DUCROCQ *et al.* (1988a), DUCROCQ (1994), RICARD Y FOURNET-HANOCQ (1997). También se han sugerido técnicas de Monte-Carlo (CLAYTON, 1991; KORSGAARD *et al.*, 1998); pero su uso para grandes masas de datos con modelo complejos (v.gr., con covariables dependientes del tiempo) podría ser computacionalmente costoso. De nuevo en un contexto bayesiano, se ha propuesto una aproximación laplaciana para la densidad marginal posterior de los hiperparámetros de la distribución de v o s (DUCROCQ y CASELLA, 1996). La validez de esta técnica se ha mostrado vía simulación. La característica atractiva es su amplia aplicabilidad a grandes masas de datos, en situaciones en las que la función de riesgo base es arbitraria (modelos de COX) o paramétrica (v.gr., modelos de WEIBULL), en que los términos de fragilidad están correlacionados según una matriz de relaciones conocida, o en modelos más generales con covariables estratificadas y/o dependientes del tiempo.

El Survival Kit

Como se ha indicado previamente, la intención primera del Survival Kit (DUCROCQ y SÖLKNER, 1984, 1998b) era llenar el vacío de software disponible para los mejoradores animales que tienden a usar masas de datos extremadamente grandes y quieren estimar efectos aleatorios. Los datos pueden ser analizados asumiendo modelos de COX, modelos Weibull o modelos con datos agrupados (en caso de datos discretos, PRENTICE y GLOECKLER, 1978). Estos modelos pueden incluir covariables continuas o discretas (posiblemente dependientes del tiempo) y funciones base separadas. Los efectos aleatorios pueden distribuirse de forma normal, normal multivariante (con matriz de parentesco) o como una función log-gamma. Se calcula el test del cociente de verosimilitudes y las curvas de supervivencia esperadas. Se incluye la estimación de los hiperparáme-

tros (v.gr., varianza genética) de los efectos aleatorios. Para aplicaciones de gran tamaño, v.gr., las evaluaciones genéticas nacionales de rutina, se ha escrito una versión del programa Weibull utilizando subrutinas C de dominio público, con objeto de limitar las operaciones de I/O comprimiendo los ficheros de datos. Se ha demostrado la factibilidad en aplicaciones a gran escala utilizando el Survival Kit por el éxito en las estimaciones de los parámetros genéticos para duración de vida productiva de vacas lecheras con más de 600.000 registros (DUCROCQ y SÖLKNER, 1998a) y la evaluación genética rutinaria para el mismo carácter en la raza Holstein en Francia que incluye unos 7,5 millones de vacas.

El Survival Kit es de libre disposición para su uso no comercial. Puede obtenerse a través de los autores (email: ducrocq@dga.jouy.inra.fr y soelkner@mail.boku.ac.at) o recuperándolo junto con el manual del usuario y algunos ejemplos en http://www.boku.ac.at/nuwi/popgen. Ha sido testado en PC (con compilador Fortran de Lahey) y sobre máquinas UNIX.

Ejemplos

Para ilustrar el uso potencial de los análisis de supervivencia, apuntamos brevemente dos aplicaciones: la implementación de una evaluación genética rutinaria de toros lecheros basada en la duración de la vida productiva de sus hijas y un análisis del efecto de la aparición de una enfermedad sobre el desecho.

Evaluación de padres

El primer intento para evaluar toros por longevidad de sus hijas utilizando técnicas de análisis de supervivencia se debe a SMITH (SMITH, 1993; SMITH y QUAAS, 1984). Se empleó un modelo de Cox estratificado por año de nacimiento de la vaca y con un efecto de grupo genético y un efecto de rebaño como únicos efectos fijos. La varianza de padre se obtuvo por un procedimiento ad hoc. Una extensión de este modelo ha sido utilizada en Austria (EGGER-DANNER, 1993), Alemania y Suiza. Los pasos para la actual evaluación genética fueron:

- El uso de un modelo de COX con covariables dependientes del tiempo como rebaño-año, estado de lactación y número de lactación para el análisis de duración de la vida productiva, en lugar de edad al desecho (DUCROCQ, 1987; DUCROCQ *et al.*, 1988a). La duración de la vida productiva se definía como el número de días entre el primer parto y la última fecha registrada.

- La demostración de que la función de riesgo base no paramétrica resultante podría ser aproximada precisamente por una función de riesgo de Weibull. Esta observación simplificaba los cálculos de manera considerable.

- La ilustración de la posibilidad de un enfoque de la longevidad funcional incluyendo en el modelo de Weibull una corrección dependiente del tiempo para desviación de producción de leche dentro de rebaño (DUCROCQ *et al.*, 1988a).

- La estimación de la varianza padre asumiendo una distribución log-gamma e integrando el efecto del padre fuera de la densidad posterior conjunta de todos los parámetros (DUCROCQ *et al.*, 1988b).

- Una mejor modelización de la parte ambiental, mostrando la influencia de la variación del tamaño del rebaño sobre el desecho, e incluyendo un efecto aleatorio rebaño-año-estación (DUCROCQ, 1994).

• La implementación de una evaluación genética de rutina de los toros lecheros en Francia, considerando solamente la longevidad funcional (corregida por producción). Los primeros resultados se obtuvieron en junio de 1997, y en la actualidad se realizan dos evaluaciones por año. El modelo es muy similar al resultante de un estudio preliminar realizado en vacas del Noreste de los Estados Unidos (datos sin publicar). Incluye los siguientes efectos fijos dependientes del tiempo: número de lactación x estado de lactación, año-estación, tamaño del rebaño x variación de este tamaño, desviación dentro de rebaño de la producción de leche y de los porcentajes de grasa y proteína. Además se incluyen dos efectos aleatorios: un efecto rebaño-año-estación distribuido como log-gamma y un efecto padre y abuelo materno distribuido como una normal multivariante (DUCROCQ y SÖLKNER, 1998a; DUCROCQ, 1999b). Modelos de evaluación similares han sido (o están siendo) implementados en Alemania, Suiza (VUKASINOVIC *et al.*, 1999), Austria, Holanda e Italia y están en estudio en Dinamarca, Suecia, Noruega, Canadá y Nueva Zelanda (no siempre para aplicaciones de rutina).

La complejidad del modelo se justifica por la necesidad de describir con precisión los cambios en las políticas de eliminación a lo largo del tiempo. Por ejemplo, aparece claramente que las vacas de rebaños en contracción o acabando su lactación al final del período de cuota (primer trimestre del calendario) tienen un mayor riesgo que las vacas en rebaños en expansión o que acaban su lactación en verano u otoño (DUCROCQ, 1994). La no consideración de estos efectos descuidaría una parte ambiental importante, reduciría nuestra capacidad para detectar diferencias genéticas entre machos y sesgaría su evaluación genética.

No es posible describir aquí en detalle todos los resultados obtenidos, aunque el cuadro 1 y las figuras 4 y 5 pueden dar una ilustración somera. El cuadro 1 muestra las principales características de los ficheros utilizados para la estimación de los parámetros genéticos y para la evaluación genética de rutina de las tres principales razas de Francia. Estos ficheros incluyen el número de vacas lecheras entre el primero de diciembre de 1984 y el 15 de mayo de 1997 y las hijas de los toros para inseminación. Para las vacas con primer parto previo a diciembre de 1984, sólo la parte correspondiente a lactaciones comenzadas después de esta fecha fueron utilizadas (registros truncados por la izquierda). Vacas con su última

Cuadro 1
Características de los ficheros de datos y resultados para las tres principales razas francesas

	Montbéliarde[1,2]	Normande[1,2]	Holstein[1]	Holstein[2]
Registros (total)	668.074	762.910	629.716	6.399.885
Censurados (%)	36,0	31,4	36,2	32,0
Truncados por la izquierda (%)	12,0	11,0	8,9	12,7
Padres	3.354	4.175	2.937	17.777
h^2 (%, escala logarítmica)	9,5	7,9	10,7	
h^2 (%, escala original)	14,1	11,8	15,9	

(1) Para la estimación de parámetros genéticos.
(2) Para la evaluación genética.

fecha de control 5 meses antes del tiempo de evaluación se consideraron aún vivas y sus registros fueron tratados como datos censurados.

La noción de heredabilidad en modelos mixtos de supervivencia merece un comentario. Como el modelo de Weibull es un caso particular de un modelo log-lineal (con una distribución de valores extremos para los residuos), la definición usual de heredabilidad (fracción de la varianza total de origen genético) puede ser aplicada en la escala log-lineal. Sin embargo, la heredabilidad en la escala log-lineal es de poca utilidad para el cálculo aproximado de los coeficientes de fiabilidad usando las fórmulas clásicas de los índices de selección, por ejemplo con una prueba de descendencia (DUCROCQ, 1999b). En cambio, se ha propuesto la definición de una heredabilidad aproximada en la escala observada y utilizarla en la fórmula del índice de selección junto con el número de descendientes no censurados (no con el número total de la progenie). De esta manera, se observó una coincidencia entre la fiabilidad realizada y predicha (DUCROCQ, 1999b). Pero la validez de este enfoque ha sido cuestionada por KORSGAARD *et al.* (1999). Un desarrollo más cuidadoso ha sido recientemente desarrollado y validado por YADZI *et al.* (2000b). Las nuevas fórmulas para la fiabilidad de la evaluación con un test de progenie y para la heredabilidad en la escala observada son

$$\text{Rel} = \frac{n_u}{n_u + (4 - h_0^2)/h_0^2} \quad \text{y} \quad h_0^2 = \frac{4\sigma_s^2}{\sigma_s^2 + 1}$$

donde n_u es el número de descendientes no censurados y σ_s^2 es la varianza de padre obtenida con el análisis de supervivencia. Se puede utilizar la heredabilidad equivalente

$$h_e^2 = \frac{4\sigma_s^2}{\sigma_s^2 + 1/p}$$

en la formula habitual

$$\text{Rel} = \frac{n}{n + (4 - h_e^2)/h_e^2}$$

donde n es el número total de descendientes del macho q y p es la proporción de la progenie que falló.

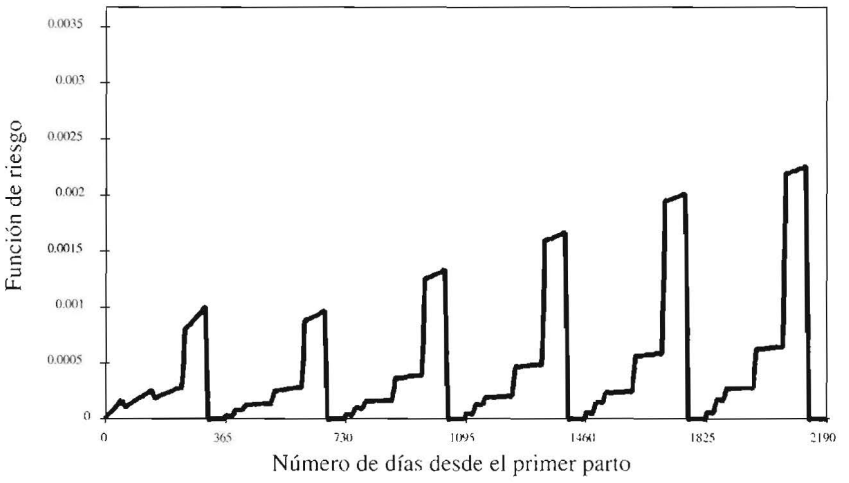


Figura 4. Cambios en el riesgo asociados con el número y el estado de lactación (en la raza Normanda, asumiendo un intervalo entre partos constante de 365 días)

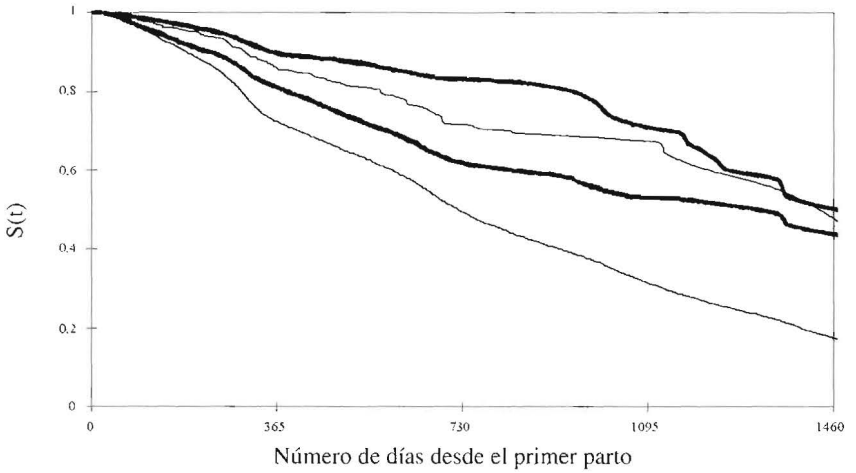


Figure 5. Curvas de supervivencia de Kaplan-Meier de las hijas de 4 toros probados de la raza normanda

En la figura 4 aparece la función de riesgo estimada para una hipotética vaca con un intervalo regular entre partos de 365 días y una lactación de 305 días. Se muestra la importancia de considerar el estado de lactación: las vacas tienen muy bajo riesgo de eliminación al comienzo de la lactación (en parte porque la última fecha de control que se gasta para calcular la duración de la vida productiva suele provenir de la lactación previa cuando la vaca es eliminada en una fase temprana después del parto). El riesgo de desecho aumenta ligeramente, y sufre un fuerte aumento al final de la lactación cuando la producción de leche es menor, cuando es mayor el valor de la canal y cuando se sabe que está preñada. La función de riesgo es similar durante las primeras dos lactaciones, aunque se incrementa en los partos posteriores (en la raza Normanda, y tras corrección por producción).

En la figura 5 aparecen las curvas brutas de supervivencia (Kaplan-Meier) de las hijas de 4 toros normandos probados. Todos tenían altos valores genéticos estimados para caracteres de producción tras las pruebas de descendencia. Sus valores de cría estimados para duración de la vida productiva corregidos por caracteres de producción fueron +2,6, +1,3, 0,3, -2,2 desviaciones estándar genéticas (entre 890 y 15.082 datos por animal). El tercer toro tenía un valor genético estimado muy bajo para conformación de ubres y para velocidad de ordeño. El cuarto tenía un EBV muy bajo para número de células somáticas. Pero el segundo también tenía bajos EBV para tipo y velocidad de ordeño, lo que condujo a que no se utilizara como padre de sementales. Las enormes diferencias observadas en supervivencia entre las hijas de estos toros (entre el 49% y el 89% de hijas vivas dos años tras su primer parto) subrayan la gran variabilidad genética existente para duración de la vida productiva y el hecho de que los EBV de la longevidad funcional proporcionan una información extra a los mejoradores, complementaria a los EBV de otros caracteres funcionales.

A pesar de sus atractivas características, hemos de aceptar que las evaluaciones de la longevidad funcional tienen todavía un impacto limitado en los esquemas existentes de mejora genética de vacuno lechero: se dispone de EBV con fiabilidades aceptables demasiado tarde, a menudo un año más tarde de que se hayan publicado los EBV de otros caracteres económicamente importantes y de que los mejores machos hayan sido ampliamente utilizados sobre la población. Para solventar este problema, la información directa sobre la longevidad tiene que ser enriquecida con información indirecta a través de predictores tempranos, v.gr., caracteres de tipo y/o número de células somáticas. Esto se está realizando en algunos países (v.gr., EEUU, Canadá, Holanda) y en otros está en estudio. El reto es decidir cuáles de los caracteres de tipo son realmente buenos predictores de la longevidad y cómo estimar de forma apropiada la correlación genética entre caracteres de tipo y longevidad. LARROQUE y DUCROCQ (2000) indican que estas cuestiones no tienen respuesta trivial, especialmente cuando en algunas subpoblaciones el desecho voluntario no está aparentemente relacionada con la funcionalidad (v.gr., para animales registrados).

Efecto de la enfermedad en la longevidad

El segundo ejemplo pertenece al campo de la epidemiología veterinaria. Las enfermedades tienen efecto sobre el desecho de las vacas lecheras. Obviamente, el momento y la severidad de una enfermedad podrían influenciar la decisión de eliminar o no una vaca y cuando eliminarla. Habitualmente, la aparición de una enfermedad ha sido tratada como una variable 0/1 (sana /enferma). Por tanto, se ha asumido de forma implícita que el riesgo de ser eliminado es el mismo durante toda la lactación, v.gr., el día en que la vaca manifiesta una enfermedad, un mes después o incluso un mes antes. Esta última característica es absurda. La figura 6a (tomada de GRÖHN *et al.* 1997) muestra las curvas de supervivencia utilizando las estimas obtenidas tras el análisis de 7.523 vacas Holstein del estado de Nueva York. La variable tiempo de fallo analizada es el número de días desde el último parto hasta el desecho y la enfermedad de interés era mastitis. Si tratamos la aparición de mastitis como una variable independiente del tiempo, nos conduce a la misma curva de supervivencia ya presente mastitis el día 1 o el día 300.

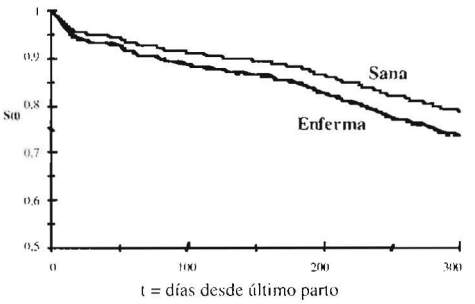


Figura 6a. Curvas de supervivencia esperada para vacas sanas y enfermas cuando la mastitis se considera como variable no dependiente del tiempo

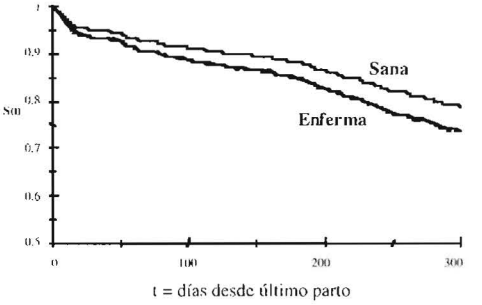


Figura 6b. Curvas de supervivencia esperada cuando la mastitis se considera como variable dependiente del tiempo

El uso de una variable enfermedad cuyos valores varían con el tiempo (0 hasta que la vaca enferma y 1 después) da como resultado una modelización más razonable del efecto de la enfermedad, como queda ilustrado en la figura 6b. La mejor adecuación del modelo también puede comprobarse con un test de cociente de verosimilitudes (véase GRÖHN *et al.*, 1997).

Es posible considerar (y validar) modelos mucho más complejos, v.gr., incluyendo la interacción entre cuando aparece la enfermedad y cuando es realmente observado el efecto sobre el desecho. Implementaciones prácticas de análisis de supervivencia en epidemiología utilizando efectos dependientes del tiempo pueden encontrarse en BEAUDEAU *et al.* (1995), GRÖHN *et al.* (1998) o NEERHOF *et al.* (2000) y una discusión sobre el tema en BEAUDEAU *et al.* (2000). Otras aplicaciones epidemiológicas incluyen estudios del efecto del conteo de células somáticas sobre aparición de mastitis en vacuno lechero (RUPP y BOICHARD, 2000) o sobre el comienzo del scrapie en ovino (ELSEN *et al.*, 1999).

Conclusión y perspectivas

Ojalá que estos ejemplos ilustren claramente algunas de las atractivas características de los análisis de supervivencia, en particular: a) toda la información se utiliza eficientemente, ya se trate de un registro observado, censurado o truncado; b) algunos de los efectos que inciden sobre la tasa de fallo varían con el tiempo. La inclusión de covariables dependientes del tiempo en un modelo de análisis de la supervivencia ofrece una formulación conveniente para tener en cuenta estas variaciones.

Debido a la importancia de la salud y el bienestar animal, la cría sostenible y unos objetivos de selección más equilibrados van a jugar progresivamente un papel central en la mejora animal. Se precisan herramientas estadísticas para estimar los componentes genéticos de la aptitud y la funcionalidad en los animales domésticos. Por tanto, no hay duda de que el análisis de supervivencia tendrá un lugar prominente entre los mejoradores animales, no sólo en la mejora del vacuno de leche donde ya se ha comenzado, sino también en otras especies. Se han publicado aplicaciones en porcino (BRANDT *et al.*, 1998; YADZI *et al.*, 2000a), en caballo (RICARD y FOURNET, 1997; WALLIN *et al.*, 2000), en avicultura (DUCROCQ *et al.*, 2000), en peces (HENRYON y BERG, 1999) y hay investigaciones en curso en vacuno de carne, en ovino y en cunicultura.

Además, una estrategia general para el análisis de datos de supervivencia en mejora animal y epidemiología puede proponerse: primero, un análisis preliminar realizado en base a un modelo de COX. Si la asunción de riesgos proporcionales no es satisfactoria, pueden utilizarse la estratificación y/o el uso de covariables dependientes del tiempo para obtener un mejor ajuste. La computación no paramétrica de la distribución de riesgo base conduce a un test gráfico indicando si es adecuado un modelo de Weibull, lo que permite reducciones en el tiempo de computación. Puede realizar el test del cociente de verosimilitudes como ayuda para la determinación de los modelos más relevantes. Si se precisa un modelo mixto, los parámetros de varianza de los efectos aleatorios son estimados a partir de su densidad marginal posterior. Asumiendo que estos componentes son conocidos, se procede a la estimación de los efectos fijos y aleatorios (y los parámetros de Weibull en el caso de un modelo de Weibull) y se estiman

las curvas de supervivencia para los animales de referencia. Estas últimas ayudan a detectar la influencia de cada variable sobre el tiempo de fallo, especialmente cuando son dependientes del tiempo (GRÖHN *et al.*, 1997). El Survival Kit se ha escrito específicamente considerando esta estrategia de análisis. Su aplicación en el ejemplo de las gallinas ponedoras es una simple pero completa ilustración.

Entre los temas importantes de investigación para el futuro estarán el desarrollo del análisis conjunto (multivariante) de algunas medidas de longevidad (tales como tiempo de fallo antes y después de que la gallina sea alojada, como en la figura 1) o de una medida de la longevidad y otros caracteres de importancia económica (v.gr., producción de leche) o predictores tempranos de la longevidad (caracteres de tipo en vacuno lechero). Este análisis multivariante es un requisito para estimar apropiadamente las relaciones genéticas y ambientales entre caracteres y su óptima incorporación a los programas de mejora existentes o futuros. Las estrategias para abordar estos análisis multivariantes han sido implementadas en un contexto gausiano (KORSGAARD *et al.*, 1999) o sugeridas en un enfoque estándar de análisis de la supervivencia (DUCROCQ, 1999c), pero su exactitud debe ser investigada en la práctica. También es necesario poder aplicarlo a grandes bases de datos, lo que implica el desarrollo paralelo de software apropiado.

Como se ha indicado previamente, muchos de los tests de bondad de ajuste son todavía gráficos. Hay una necesidad precisa de más herramientas formales a diferentes niveles, v.gr., el contraste de la asunción de censura aleatoria, la elección de formas paramétricas para la función base, el chequeo de la especificación errónea de las relaciones entre covariables y tiempo de fallo, la necesidad de incluir otras covariables en el modelo, la detección de la sobredispersión, la medida de la discrepancia entre curvas de supervivencia predichas y observadas, la medida de la fracción de variación explicada por el modelo, etc. Algunas ya han sido propuestas (KLEIN y MOESCHBERGER, 1997) pero a menudo carecen de viabilidad computacional para grandes bases de datos o para modelos complejos con covariables dependientes del tiempo. La existencia de varias (>6) definiciones diferentes de los residuos da una idea de que el problema no está resuelto. Los estadísticos están trabajando intensamente en estas direcciones.

Por último, otra futura área de investigación va a ser la extensión de los modelos de supervivencia a otros modelos genéticos, v.gr., incluyendo información de marcadores o de genes mayores.

Bibliografía

BEAUDEAU F., DUCROCQ V., FOURICHON C., SEEGERS H., 1995. Effect of disease on length of productive life of French Holstein cows assed by survival analysis. J. Dairy Sci., 78: 103-117.

BEAUDEAU F., SEEGERS H., DUCROCQ V., FOURICHON C., BAREILLE N., 2000. Effect of health disorders on culling in dairy cows: a review and critical discussion. Ann. Zootechnie, (en prensa).

BRANDT H., VON BREVERN N., GLODEK P., 1998. Factors affecting survival rate of crossbred sows in weaner production. Livest. Prod. Sci., 57: 127-135.

CLAYTON D.G., 1991. A Monte-Carlo method for Bayesian inference in frailty models. Biometrics, 47:467-485.

CLAYTON D.G., CUZICK J., 1985. Multivariate generalizations of the proportional hazards model. J. Royal Stat. Soc., Series A, 148:82-117.

COX D.R., 1972. Regression models and life table. J. Royal Stat. Soc., Series B, 34:187-220.

COX D.R., OAKES D., 1984. Analysis of survival data. Chapman and Hall, London, UK.

DUCROCQ V., 1987. An analysis of length of productive life in dairy cattle. Tesis Doctoral, Cornell University, Ithaca, NY, USA.

DUCROCQ V., 1994. Statistical analysis of length of productive life for dairy cows of the Normande breed. J. Dairy Sci., 77:855-866.

DUCROCQ V., 1999a. Extension of survival analysis to discrete measures of longevity. Fourth International Workshop on Genetic Improvement of Functional Traits in Cattle : Longevity, Jouy-en-Josas, May 9-11, 1999, Interbull Bulletin N.° 21, 41-47. Uppsala, Sweden.

DUCROCQ V., 1999b. Two years of experience with the French genetic evaluation of dairy bulls on production-adjusted longevity of their daughters. Fourth International Workshop on Genetic Improvement of Functional Traits in Cattle : Longevity, Jouy-en-Josas, May 9-11, 1999, Interbull Bulletin N.° 21, 60-67. Uppsala, Sweden.

DUCROCQ V., 1999c. Topics that may deserve further attention in survival analysis applied to dairy cattle breeding : some suggestions. Fourth International Workshop on Genetic Improvement of Functional Traits in Cattle: Longevity, Jouy-en-Josas, May 9-11, 1999, Interbull Bulletin N.° 21, 181-189. Uppsala, Sweden.

DUCROCQ V., BESBES B., PROTAIS M., 2000. Genetic improvement of laying hens viability using survival analysis. Genet. Sel. Evol., 32, 23-40.

DUCROCQ V., CASELLA G., 1996. A Bayesian analysis of mixed survival models. Genet. Sel. Evol., 28: 505-529.

DUCROCQ V., QUAAS R.L., POLLAK E. J., CASELLA G., 1988a. Length of productive life of dairy cows. I. Justification of a Weibull model. J. Dairy Sci., 71:3061-3070.

DUCROCQ V., QUAAS R.L., POLLAK E. J., CASELLA G., 1988b. Length of productive life of dairy cows. II. variance component estimation and sire evaluation. J. Dairy Sci., 71:3071-3079.

DUCROCQ V., SÖLKNER J., 1994. "The Survival Kit", a FORTRAN package for the analysis of survival data. In: 5th World Cong. Genet. Appl. Livest. Prod., Volume 22, pages 51-52. Dep. Anim. Poultry Sci., Univ. Of Guelph, Guelph, Ontario, Canada.

DUCROCQ V., SÖLKNER J., 1998a. Implementation of a routine breeding value evaluation for longevity of dairy cows using survival analysis techniques. In : 6th World Cong. on Genet. Appl. Livest. Prod., January 11-16, University of New-England, Armidale, Australia, Vol 23, 359-362.

DUCROCQ V., SÖLKNER J., 1998b. "The Survival Kit - V3.0", a package for large analyses of survival data. In : 6th World Cong. on Genet. Appl. Livest. Prod., January 11-16, University of New-England, Armidale, Australia Vol 27, 447-450.

EGGER-DANNER C., 1993. Zuchtwerschätzung für Merkmale der Langlebigkeit beim Rind mit Methoden des Lebensdaueranalyse. Tesis Doctoral, University of Agriculture, Vienna, Austria.

ELSEN J.M., AMIGUES Y., SCHELCHER F., DUCROCQ V., ANDREOLETTI O., EYCHENNE F., VU TIEN KHANG J., POIVEY J.P., LANTIER F., LAPLANCHE J.L., 1999. Genetic susceptibility and transmission factors in scrapie: detailed analysis of an epidemic in a closed flock of Romanov. Archives of Virology, 144, 431-445.

FOLLMANN D.A., GOLDBERG M.S., 1988. Distinguishing heterogeneity from decreasing hazard rates. Technometrics, 30:38-396.

GRÖHN Y.T., DUCROCQ V., HERTL J.A., 1997. Modeling the effect of a disease on culling : an illustration of the use of time-dependent covariates in survival analysis. J. Dairy Sci., 80: 1755-1766.

GRÖHN Y.T., EICKER S. W., DUCROCQ V., HERTL J.A., 1998. The effect of disease on culling in New-York State Holstein dairy cows. J. Dairy Sci., 81, 966-978.

HENRYON M., BERG P., 1999. Genetic variation for traits of commercial importance exists in Danish rainbow trout. 50th Annual Meeting of the EAAP, August 23-26, Zurich, Switzerland, 5:36 (abstract).

HOUGAARD P., 1986. A class of multivariate failure time distributions. Biometrika, 73:671-678.

KALBFLEISCH J.D., PRENTICE R.L., 1980. The statistical analysis of failure time data. John Wiley and sons, New-York, USA.

KAPLAN E.L., MEIER P., 1958. Nonparametric estimation from incomplete observations. J. Amer. Stat. Ass., 53 :457-481.

KLEIN J.P., 1992. Semiparametric estimation of random effects using the Cox model based on the EM algorithm. Biometrics, 48:795 - 806.

KLEIN J.P., MOESCHBERGER M., 1997. Survival analysis. Springer-Verlag, New-York, USA.

KORSGAARD I.R., ANDERSEN A.H., JENSEN J., 1999. Discussion of heritability of survival traits. Fourth International Workshop on Genetic Improvement of Functional Traits in Cattle: Longevity, Jouy-en-Josas, May 9-11, 1999, Interbull Bulletin N.° 21, Uppsala, Sweden.

KORSGAARD I.R., LUND M.S., MADSEN P., JENSEN J., SORENSEN D., GIANOLA D., 1999. Multivariate Bayesian analysis of Gaussian, righ-censored Gaussian, ordered categorical and binary traits using Gibbs sampling. Proc. Computational Cattle Breeding. March 18-20, Tuusula, Finland, Interbull Bulletin N.° 20, 79-92.

KORSGAARD I.R., MADSEN P., JENSEN J., 1998. Bayesian inference in the semiparametric lognormal frailty model using Gibbs sampling. Genet. Sel. Evol., 30:241-256.

LARROQUE H., DUCROCQ V., 2000. Relationship between type and longevity in the Holstein breed. Genet. Sel. Evol., (en prensa).

NEERHOF H.J., MADSEN P., DUCROCQ V., VOLLEMA A. R., JENSEN J., KORSGAARD I.R., 2000. Journal of Dairy Science, 83, 1064-1071.

PRENTICE R.L., GLOECKER L.A., 1978. Regression analysis of grouped survival data with application to breast cancer data. Biometrics, 34 : 57-67.

RICARD A., FOURNEL-HANOCQ F., 1997. Analysis of factors affecting length of competitive life of jumping horses. Genet. Sel. Evol., 29, 251-267.

RUPP R., BOICHARD D., 2000. Relationship of early first lactation somatic cell count with risk of subsequent first clinical mastitis. Livest. Prod. Sci., 62: 169-180.

SMITH S., 1983. The extension of failure time analysis to problems of animal breeding. Tesis Doctoral, Cornell University, Ithaca, NY, USA.

SMITH S.P., QUAAS R.L., 1984. Productive life span of bull progeny groups: failure time analysis. J. Dairy Sci., 67:2999-3007.

VAUPEL J., MANTON K.G., STALLARD E., 1979. The impact of heterogeneity in individual frailty and the dynamics of mortality. Demography, 16:439-454.

VUKASONOVIC N., MOLL J., KÜNZI N., 1999. Genetic evaluation for length of productive life with censored records. J. Dairy Sci., 82: 2178-2185.

YAZDI M.H., RYDHMER L., RINGMAR-CEDERBERG E., LUNDEHEIM N., JOHANSSON K., 2000a. Genetic study of longevity in Swedish Landrace sows. Livest. Prod. Sci. , 63, 255-264.

YAZDI M.H., VISSCHER P.M., V. DUCROCQ, R. THOMPSON, 2000b. Genetic parameters and response to selection in proportional hazard models. J. Dairy Sci. (abstract), Supp. 1

WALLIN L., STRANDBERG E., PHILIPSSON J., DALIN G., 2000. Estimates of longevity and causes of culling and death in Swedish warmblood and coldblood horses. Livest. Prod. Sci., 63, 275-289.

ITEA (2000). Vol. **96A N.° 3,** 254-260

POLIMORFISMO *Hha*I DEL GEN DE LA HORMONA DE CRECIMIENTO EN CERDOS IBÉRICOS: RELACIÓN CON CARACTERES PRODUCTIVOS

A. Fernández, C. Óvilo, C. Rodríguez, L. Silió

Dpto. Mejora Genética y Biotecnología SGIT-INIA, Ctra. Coruña, km 7, 28040 Madrid, España

RESUMEN

En el presente trabajo se estudia un polimorfismo del gen de la hormona de crecimiento porcina (*GH*) en distintas poblaciones de cerdos de las razas Ibérica y Duroc y se analiza la asociación de variantes alélicas con caracteres de crecimiento y composición de la canal, en una muestra de 199 cerdos ibéricos de la estirpe Torbiscal. Esta población está compuesta de dos líneas, una línea de conservación y otra seleccionada para crecimiento magro, que presentan diferencias significativas en algunos de estos caracteres productivos. El polimorfismo utilizado es del tipo PCR-RFLP (*Hha*I) y se localiza en el segundo exón del gen. El estudio de asociación del polimorfismo se realizó utilizando un modelo animal en el que el genotipo del gen *GH* se incluye como efecto fijo. A pesar de las evidencias previas de asociación de este polimorfismo con distintos caracteres de interés económico y de la diferencia en las frecuencias alélicas de las dos líneas estudiadas, no se detecta asociación de este marcador con ninguno de los caracteres analizados en el presente trabajo.

Palabras clave: Cerdo, Gen de la hormona de crecimiento, Gen candidato, Caracteres de crecimiento y composición corporal.

SUMMARY

*Hha*I POLYMORPHISM OF THE GROWTH HORMONE GENE IN IBERIAN PIGS: RELATION WITH PRODUCTIVE TRAITS

In this work we analyze a polymorphism in the porcine growth hormone gene (*GH*) in different populations of Iberian and Duroc pigs, and we study the association of the different alleles with growth and carcass traits in a sample of 199 Iberian pigs from the Torbiscal strain. This population includes two lines, the conservation line and a selected one, which present significant differences in some of the productive traits. The polymorphism used is a PCR-RFLP (*Hha*I) localized in the gene's second exon. The association study was performed using an animal model in which the genotype was included as a fixed effect. Instead of the previous association evidences of this polymorphism with several economic traits and the difference in the allelic frequencies between the studied lines, we could not detect any association of this marker with the traits analyzed.

Key words: Pig, Growth hormone gene, Candidate gene, Growth and carcass traits.

Introducción

Los estudios de asociación de genes candidatos son un método adecuado para la identificación de genes relacionados con caracteres productivos, y son utilizados actualmente como alternativa o complemento a los experimentos de detección de QTLs mediante el barrido genómico con marcadores anónimos. Estos dos tipos de estudios permiten la aplicación de la tecnología del ADN al conocimiento de la base genética de los caracteres de interés económico. Por otro lado, la aproximación del gen candidato facilita el desarrollo, en caso de confirmarse asociaciones significativas entre el gen y dichos caracteres, de kits moleculares que harían más asequible el empleo de esta información molecular en la mejora genética de animales domésticos.

El gen de la hormona de crecimiento (*GH*), se suele considerar buen candidato para ciertos caracteres productivos, habiendo sido objeto de múltiples estudios de asociación, debido a la relación directa que esta hormona muestra con el crecimiento y la composición corporal, ya que es el principal factor endocrino en la regulación del desarrollo muscular. El gen *GH* esta localizado en cerdos en el brazo corto del cromosoma 12, constituido por 5 exones y con un transcrito total de 1,7 Kb. En los últimos años se han desarrollado múltiples trabajos que han permitido identificar numerosos polimorfismos en este gen, principalmente del tipo PCR-RFLP, en la región flanqueante 5', segundo exón y regiones intrónicas (DWORAK *et al.*, 1997; JIANG *et al.*, 1996; KUCIEL *et al.*, 1992). En muchos de estos trabajos se ha tratado de encontrar alguna asociación entre los distintos polimorfismos y caracteres de composición de la canal y engrasamiento con resultados, en general, contradictorios. Según CASAS-CARRILLO *et al.* (1997) el efecto de la sustitución de una base en la región promotora y en el segundo exón del gen sobre el crecimiento y la composición de la canal en cerdo es trivial. NIELSEN *et al.* (1995) sugieren la existencia de diferencias en la actividad transcripcional entre distintas variantes del gen *GH*, afectando a caracteres de crecimiento en cerdo. DWORAK *et al.* (1996) encuentran cierta asociación con la ganancia diaria en peso, KNORR *et al.* (1997) concluyen que el gen *GH* esta relacionado con la variación observada en engrasamiento y sugieren que el efecto puede atribuirse a variantes dentro del propio gen y no a QTLs ligados.

En este trabajo se presentan los resultados obtenidos en el estudio de la asociación del polimorfismo PCR-RFLP, *Hha*I descrito por LARSEN y NIELSEN (1993) con caracteres de crecimiento y composición de la canal en poblaciones de cerdo de la raza Ibérica.

Material y métodos

El polimorfismo fue analizado en 428 animales de los cuales, 350 corresponden a la estirpe Torbiscal, donde se cuenta con una línea seleccionada para mejor crecimiento magro en relación con la línea control (SILIO *et al.*, 1997), 30 animales de la estirpe Guadyerbas (TORO *et al.*, 2000) y 48 animales inscritos en el Libro Genealógico y correspondientes a otras estirpes de la raza Ibérica. También fueron analizados 44 animales pertenecientes a distintos orígenes de la raza Duroc.

Los genotipos fueron identificados por PCR-RFLP. Para la reacción de PCR se utilizaron cebadores específicos (LARSEN y NIELSEN, 1992), que permiten la amplifica-

ción de un fragmento de 605 pb, comprendido entre las posiciones +61 y +665, incluyendo el primer y segundo exón del gen de la hormona de crecimiento porcino.

El protocolo de genotipado se realizó como se describe en el trabajo de OVILO *et al.* (1999). Los genotipos fueron identificados por la aparición de fragmentos con los siguientes tamaños, estimados a partir de la secuencia publicada del gen *GH* (Genbank, n.º de acceso: M17704): alelo C1: 605 pb; alelo C2: 496 y 109 pb; alelo C3: 447 y 158 pb; alelo C4: 447, 109 y 49 pb.

Los análisis de asociación del polimorfismo con los caracteres medidos se realizaron sobre un subgrupo de 199 animales, pertenecientes a ambas líneas de la estirpe Torbiscal: 148 Conservación (Torbiscal-C) y 53 Selección (Torbiscal-S). Los caracteres analizados en el estudio fueron: crecimiento en recría (GMD1), medido como ganancia de peso diaria desde el destete hasta el inicio del periodo de ceba; Espesor de grasa dorsal, medido a los 100 kg; Crecimiento en ceba (GMD2), medido como la ganancia de peso diaria desde que finaliza el periodo de recría hasta el sacrificio y porcentajes sobre peso de la canal de jamones, paletas y lomos.

El análisis estadístico de asociación fue realizado con el programa PEST, para estimar el efecto de cada genotipo sobre los caracteres medidos. El modelo lineal usado incluye como efectos fijos la línea (dos niveles), año (tres niveles) y genotipos del polimorfismo *Hha*I (tres niveles), y en el análisis de las piezas nobles el peso de la canal como covariable. Para cada carácter analizado fueron estimados dos contrastes: la diferencia entre los genotipos homocigotos ($2a$ = C3C3-C4C4), y el efecto dominante, estimado como la desviación de los genotipos heterocigotos de la media de los genotipos homocigotos (d = C3C4-1/2(C3C3+C4C4)).

Para realizar estos análisis se llevó a cabo la estimación previa de componentes de varianza, mediante máxima verosimilitud restringida, utilizando el programa VCE. El modelo de análisis fue similar al descrito previamente, excluyendo el efecto del polimorfismo, disponiéndose de registros productivos de 610 machos castrados (343 de la línea C y 267 de la línea S), sacrificados en los últimos seis años. El total de animales en la genealogía ascendió a 1004.

Resultados

Se analizó el polimorfismo en las distintas poblaciones de ambas razas, encontrando la presencia de sólo dos alelos (C3 y C4) de los cuatro previamente descritos, con la única excepción de un animal Duroc con genotipo C1C4. El predominio de estos alelos está en concordancia con los resultados de experimentos previos en distintas razas porcinas en las que, a pesar de aparecer los cuatro alelos descritos, presentan generalmente los alelos C1 y C2 a baja frecuencia (HANDLER *et al.*, 1996; JIANG *et al.*, 1996; KUCIEL *et al.*, 1998).

En el cuadro 1 se presentan las frecuencias relativas para cada genotipo en cada una de las poblaciones estudiadas. Se puede observar la existencia de los tres genotipos en Torbiscal-C, Torbiscal-S y Duroc, mientras que en la estirpe Guadyerbas sólo aparece representado el alelo C3. Respecto al grupo de otros animales ibéricos, al tratarse de un conjunto muy heterogéneo, formado por pequeñas poblaciones cerradas, comprende estirpes donde aparecen los dos alelos representados a distintas frecuencias, a

excepción del caso de Coronado, donde sólo aparece representado el alelo C3, en concordancia con la otra estirpe de cerdo Negro Lampiño (Guadyerbas).

Las frecuencias genotípicas difieren entre las dos líneas de la estirpe Torbiscal ($\chi^2 = 44.59$, $p < 0{,}0001$), siendo significativamente mas frecuente el alelo C3 en Torbiscal-S y el C4 en Torbiscal-C.

En el cuadro 2 se presentan la media, desviación típica y rango de variación de los caracteres registrados en los 510 animales Torbiscal con información fenotípica disponible. En el cuadro 3 figuran las estimas de los parámetros genéticos (heredabilidades y correlaciones genéticas) de estos caracteres. En el cálculo de estas estimas se ha asumido correlación nula entre los grupos de caracteres de crecimiento y rendimiento de piezas nobles.

Las estimas de cocientes de varianzas (heredabilidades y correlaciones) obtenidas para los caracteres de composición corporal son algo más elevadas que las encontradas en la literatura para este tipo de caracteres. Estos elevados valores pueden deberse a que las condiciones de manejo, sacrificio y despiece de los animales han sido relativamente homogéneas a lo largo de los seis años de control. Se han detectado diferencias genéticas entre las líneas Torbiscal (S-C), que son estadísticamente significativas para el espesor de grasa a 100 kg y el porcentaje de paletas sobre el peso de la canal.

En el cuadro 4 se indica la diferencia entre los genotipos homocigotos (*2a*) y el efecto dominante (*d*) de los genotipos encontrados sobre los caracteres de crecimiento y composición de la canal. Estos resultados corresponden a los 199 animales Torbiscal de los cuales se dispone de información fenotípica y genotípica. Ninguno de los efectos observados alcanza significación estadística.

Discusión

La administración exógena de *GH* produce en los animales un aumento en la masa muscular y una reducción del engrasamiento, así como una mejora en el índice de conversión y en la ganancia diaria de peso. Por otra parte existen estudios moleculares que relacionan ciertas variantes alélicas con diferencias en la actividad transcripcional del gen que codifica esta hormona. Esto implica

Cuadro 1
Frecuencias genotípicas para el polimorfismo *Hha*I
Table 1
Frequencies of genotypes for HhaI *polymorphism*

Población	N	C3C3	C3C4	C4C4
Torbiscal-C	229	0,280	0,563	0,157
Torbiscal-S	121	0,636	0,331	0,033
Guadyerbas	30	1	0	0
Otros ibéricos	48	0,396	0,354	0,250
Duroc	44*	0,163	0,488	0,349

(*) Se identificó un animal con genotipo C1/C4.

Cuadro 2
Datos productivos utilizados en el análisis
Table 2
Productive data used in the analysis

Caracteres	Media	DT	Rango de variación
GMD1 (g/d)	387	43	242 – 512
GMD2 (g/d)	521	74	246 – 746
Espesor grasa dorsal(mm)	26,7	4,7	14,9 – 38,5
Peso canal (kg)	126,9	11,6	90,9 – 160,4
% jamones	16,8	0,8	14,4 – 20,2
% paletas	11,4	0,6	9,9 – 13,9
% lomos	3,0	0,3	2,1 – 4,3

Cuadro 3
Heredabilidades (diagonal) y correlaciones genéticas de caracteres de crecimiento y canal
Table 3
Heredabilities (diagonal) and genetic correlations of growth and carcass traits

Carácter	GMD1	GMD2	Espesor grasa	% jamones	% paletas	% lomos
GMD1	0,48(.04)	—	—	—	—	—
GMD2	0,19(.14)	0,29(.07)	—	—	—	—
Espesor grasa	-0,34(.11)	-0,42(.18)	0,56(.09)	—	—	
% jamones	—	—	—	0,88(.09)	—	—
% paletas	—	—	—	0,73(.07)	0,47(.08)	—
% lomos	—	—	—	0,52(.08)	0,47(.08)	0,85(.08)

Errores típicos entre paréntesis.

Cuadro 4
Efectos aditivo y dominante del polimorfismo *GH HhaI* sobre caracteres de crecimiento y composición de la canal
Table 4
Additive and dominant effects of GH Hha*I polymorphism on growth and carcass traits*

Carácter	*2a*	*d*
GMD1 (g/d)	5 ± 11	5 ± 7
GMD2 (g/d)	-4 ± 18	-3 ± 11
Espesor grasa dorsal (mm)	-1,27 ± 0,95	-0,86 ± 0,57
% jamones	-0,09 ± 0,17	-0,06 ± 0,10
% paletas	-0,01 ± 0,12	0,03 ± 0,07
% lomos	0,01 ± 0,06	0,01 ± 0,04

que los distintos alelos del gen puedan ser responsables de distintos niveles plasmáticos de *GH* y por tanto también de parte de la variación observada en los caracteres de crecimiento y engrasamiento.

El polimorfismo *Hha*I utilizado en este trabajo está localizado en el segundo exón del gen y modifica la composición aminoacídica de la proteína codificada lo que hace suponer que podría modificar su actividad hormonal. Por otra parte, la diferencia significativa observada en las frecuencias alélicas en las dos líneas de la estirpe Torbiscal indicaría una posible asociación de estos alelos con los caracteres seleccionados.

Entre los diversos experimentos previos que han detectado diferencias fenotípicas en caracteres productivos asociadas a las distintas variantes moleculares del gen, varios han analizado el polimorfismo *Hha*I, siendo los resultados bastante dispares. Entre ellos, en el trabajo de DWORAK *et al.* (1996) se describe la asociación de este polimorfismo con la ganancia diaria de peso. Sin embargo, los dos caracteres de crecimiento GMD1 y GMD2, analizados en el presente trabajo no han mostrado asociación con las variantes del gen, con lo que no se ha podido confirmar el efecto descrito previamente para el primero de ellos.

KNORR *et al.* (1997) han descrito la asociación del polimorfismo analizado con distintos caracteres de engrasamiento en poblaciones procedentes del cruce de las razas Meishan x Pietrain, mientras que esta asociación no se detecta en los cruces de Jabalí con Pietrain. Estos autores sugieren que esta diferencia entre poblaciones podría deberse a que el efecto de la somatotropina porcina depende del potencial genético para deposición grasa de la raza, siendo mayor en razas muy grasas no seleccionadas. En este sentido cabría esperar también una influencia de este gen sobre la deposición grasa en la población Ibérica utilizada en este estudio. Entre los caracteres de engrasamiento que muestran asociación en los trabajos citados está el espesor de la grasa dorsal analizado también en este trabajo. Sin embargo, los resultados obtenidos en la población Ibérica no permiten confirmar la asociación del gen *GH* con este carácter.

MOSER *et al.* (1996) detectan, en cruces experimentales Jabalí x Pietrain y Meishan x Pietrain, efectos significativos del polimorfismo *Hha*I sobre distintos caracteres de composición de la canal. Sin embargo, la asociación del polimorfismo con alguno de los caracteres de composición de la canal registrados en este trabajo no ha podido confirmarse.

Por otra parte existen también trabajos que coincidiendo con nuestros resultados descartan la asociación de distintos polimorfismos en el gen, entre ellos el RFLP *Hha*I, con varios caracteres de crecimiento y composición de la canal (CASAS-CARRILLO *et al.*, 1997).

Considerando el conjunto de resultados de distintos trabajos relativos al gen *GH*, podría sugerirse que el efecto observado en algunos de los trabajos que analizan el polimorfismo *Hha*I, podría ser debido a otro polimorfismo en el mismo gen, de modo que la detección del efecto en los estudios de asociación dependiera de la fase de ligamiento con el marcador analizado y de los haplotipos presentes en cada población. Otra explicación podría ser que el efecto se debiera a otro gen o genes próximos, aunque ello parece en principio menos probable. Los numerosos experimentos realizados de mapeo de QTLs en porcino no han aportado resultados relevantes en esta región y tampoco se conocen en el cromosoma 12 otros genes candidatos para caracteres de crecimiento o composición corporal. La dificultad de asignar

un efecto fenotípico a un polimorfismo concreto detectado en un gen es un problema común de la estrategia del gen candidato, que en este caso se pone en evidencia debido a la gran cantidad de trabajos relativos a distintos polimorfismos que se han realizado sobre el gen de la hormona de crecimiento.

Por otro lado en la interpretación de los resultados de asociación de genes candidatos, también hay que tener en cuenta que en el efecto de un determinado alelo o haplotipo influye el ambiente genético y las interacciones con otros genes que también afectan a estos caracteres productivos, como es el caso de los genes relacionados con el eje somatotrópico cuyas interacciones son conocidas (NIELSEN *et al.*, 1995).

Agradecimientos

Al personal del CIA Dehesón del Encinar por su dedicación al manejo de los animales. A las industrias cárnicas INPORCASA y Chacinas de Candelario por las facilidades en la obtención de registros de despiece. A. Fernández disfruta de una beca de la Comunidad de Madrid. Proyecto financiado por el INIA (proyecto SC98-083).

Bibliografía

CASAS-CARRILLO E., PRILL-ADAMS A, PRICE S G., CLUTTER A C., KIRKPATRICK B W., 1997. Relationship of growth hormone and insuline-like growth factor-1 genotypes with growth and carcass traits in swine. Animal Genetics, 28, 88-93.

DWORAK J., MAYR B., TAXACHER A., SEIWALD G., KLEIBEL A., 1996. Growth hormone polyorphisms, growth performance and meat quality in Austrian slaughtering pigs. 25Th International Conference on Animal Genetics, ISAG (Tours, France).

DWORAK J., VRTKOVÁ I., KAHÁKOVÁ L., 1997. Genotypes of candidate genes for QTL in slaughter pigs in the Czech Republic. 49Th Annual Meeting of the EAAP(Viena)

HANDLER J., SHOMOLL F., STUR I., BREM G., SCHELLANDER K., 1996. Distribution of ApaI and CfoI polymorphism of the porcine growth-hormone (pGH) gene in two ryr 1 genotype Austrian pig breeds. J. Anim. Breed. Genet.113, 56-61.

JIANG Z.H., ROTTMANN O.J., PIRCHNER F., 1996. *Hha*I enzyme reveals genetic polimorphisms at the second exon of porcine-growth-hormone gene. J. Anim. Breed: Genet. 113, 553-558.

KNORR C., MOSER G., MÜLLER E., GELSERMANN H., 1997.Associations of *GH* gene variants with performance traits in F2 generations of European wild boar, Piétrain and Meishan pigs. Animal Genetics, 28, 124-128.

KUCIEL J., KRENKOVA L., URBAN T., 1998. Polymorphisms of HaeII and MspI growth hormone genes in commercial hybrid pigs with different RYR1 genotypes. J. Anim. Breed. Genet. 115, 397-402.

LARSEN N.J., NIELSEN V.H., 1993. ApaI and CfoI polymorphisms in the porcine growth hormone gene. Animal Genetics. 24, 71.

MOSER G., KNORR C., GELDERMANN H., BRUNSCH C., HEYDRICH C., LEUTHOLD G., 1996. Association between growth and Pit-1 genotypes with production traits in informative F2 families. 47Th Annual Meeting of the European Association for Animal Production (Lillehammer).

NIELSEN V.H., LARSEN N.J., AGERGAARD N., 1995. Association of DNA polymorphism in the growth-hormone gene with basal-plasma growth-hormone concentration and production traits in pigs. J. Anim. Breed. Genet. 112. 205-212.

ÓVILO C., FERNÁNDEZ A., 1999. Análisis de polimorfismos en el gen de la hormona de crecimiento: Estudio de la segregación de un cruce F2 entre las razas Ibérica y Landrace. ITEA, Vol. Extra 20, 294 - 296.

SILIÓ L., TORO M.A., RODRÍGUEZ C., RODRIGÁÑEZ J., 1997. Inferencias sobre cambios genéticos en una línea de cerdos ibéricos seleccionada para crecimiento magro. Invest. Agr.: Prod. Sanid. Anim. 12 (1, 2 y 3), 65-77.

TORO M.A., RODRIGÁÑEZ J., SILIÓ L., RODRÍGUEZ C., 2000. Genealogical analysis of a closed herd of black hairless iberian pigs. Conservation Biology.

ITEA (2000), Vol. **96A N.° 3,** 261-269

LA FUERZA DEL MESTIZO: ESTIMACIÓN DE LA HETEROSIS EN EL CRECIMIENTO DE CERDOS IBÉRICOS EN MONTANERA

A. Fernández, M.A. Toro, J. Rodrigáñez, M.C. Rodríguez, L. Silió

Departamento de Mejora Genética y Biotecnología, SGIT-INIA, Madrid, España

RESUMEN

A partir de 1917 registros de peso de 579 animales procedentes de los 16 cruces posibles entre cuatro estirpes de cerdo Ibérico, se ha realizado un análisis dialélico para la estima de los parámetros de cruzamiento en dos caracteres de crecimiento: peso a los 14 meses (P_{14M}) y ganancia media diaria (GMD) en ceba. Se ha comparado un modelo de efectos fijos con un modelo animal bivariante, empleando para la resolución de este último tres procedimientos diferentes: a) máxima verosimilitud restringida, b) análisis bayesianos de funciones lineales de crecimiento sin y c) con información a priori. Los errores de los parámetros estimados se subestiman con el modelo de efectos fijos y aumentan con el grado de incertidumbre atribuido a los componentes de varianza. Ello unido al uso más eficiente de la información que permiten las técnicas bayesianas, hace que las inferencias respecto a algunos parámetros se modifiquen en función del procedimiento utilizado. Los resultados mostraron una heterosis media significativa para ambos caracteres que expresados como porcentaje de la media de las estirpes puras oscilan entre 9,8 y 16% (P_{14M}) y 5,5 y 10% (GMD) para la ganancia media diaria. Estos valores confirman el interés del cruzamiento entre líneas como una herramienta simple en la mejora genética de crecimiento de cerdos Ibéricos.

Palabras clave: Cruce dialélico, Heterosis, Crecimiento, Cerdos Ibéricos.

SUMMARY

A datafile of 1917 weigth records of 579 animals from the 16 possible crosses among 4 strains of iberian pigs was analyzed in order to estimate crossbreeding parameters for two growth traits: weigth at 14 m. (P_{14M}) and average daily gain during the final fattening (GMD). A fixed model was compared with a bivariant animal model solved using three alternative methods: restricted maximum likelihood and a bayesian analysis of intercept and slope of linear growth function including or not a priori information about (co)variance components. The standard errors of the estimated parameters are underestimated with the fixed model and increase with the uncertainty of the variance components. The most efficient use of data allowed by the bayesian techniques also contributes to modify the inferences about some parameters according the used procedure. The results show a significant mean heterosis for both traits with values, expressed as percentage of the average of the pure strains, ranging from 9.8 to 16% (P_{14M}) and from 5.5 to 10% (GMD). These values validate the interest of the crossbreeding between lines as a useful tool to improve growth rates in the final fattening of Iberian pigs.

Key words: Diallel cross, Heterosis, Growth traits, Iberian pigs.

Introducción

Los diseños de cruce dialélico, en los que a partir de N parentales se obtienen los N^2 cruzamientos posibles, se han utilizado ampliamente en plantas como un procedimiento de identificar variación genética aditiva y no aditiva en caracteres de interés. Por diversas razones su empleo en animales domésticos ha sido más limitado aunque existen ejemplos incluso en vacuno de carne (COMERFORD y col., 1988). Las poblaciones ganaderas parentales no pueden considerarse muestras aleatorias de genotipos y el interés de estos diseños en mejora animal se restringe a la identificación de las combinaciones genéticas de mayor interés productivo. Asimismo, excepto en casos excepcionales de líneas consanguíneas, en general se trata de poblaciones genéticamente no homogéneas. Los diseños alélicos suponen una no asociación entre genotipos y cruces de difícil realización en animales domésticos en los que es casi inevitable disponer de datos desequilibrados. Por ello se ha recomendado el uso del modelo animal en el análisis de cruces dialélicos que presenta claras ventajas respecto a los modelos de efectos fijos. Como han indicado KOMENDER y HOESCHELE (1989), el empleo del modelo animal requiere disponer de estimas adecuadas de los componentes de varianza, cuya estimación a partir de la información obtenida en el dialélico es en ocasiones imprecisa.

El objetivo del presente trabajo es mostrar una aplicación de las técnicas bayesianas de análisis de funciones de producción (VARONA y col., 1997), que permiten el uso mas eficiente de la información disponible, a la estimación de parámetros de cruzamiento a partir de registros de crecimiento obtenidos en un cruce dialélico entre cuatro antiguas estirpes de cerdo Ibérico.

Material

Los datos analizados en el presente estudio se registraron en la piara experimental de cerdos ibéricos de "El Dehesón del Encinar" (Oropesa, Toledo). En el periodo inicial de esta piara se ensayaron cruzamientos entre sus cuatro estirpes fundadoras, *Ervideira* (1), *Campanario* (2), *Caldeira* (3) y *Puebla* (4), una parte de ellos de acuerdo con un diseño de cruce dialélico completo. En este trabajo se ha utilizado un total de 1917 pesadas, registradas durante la ceba en montanera de 579 cerdos ibéricos, 287 machos castrados y 292 hembras, nacidos en tres años consecutivos (1951-1953). La edad media de los animales al comienzo de la montanera era de 356 días con una edad y peso medio al sacrificio de 461 días y 155,8 kg. El número de pesadas utilizado osciló de 1 a 4 registros por individuo. En el cuadro 1 se muestra el número de animales con datos correspondientes a cada uno de los 16 tipos de cruces así como el número de verracos y cerdas madres utilizados como reproductores. La distribución de las progenies de estos reproductores por tipo de cruzamiento esta lejos de ser equilibrada. El análisis de la correspondiente tabla de contingencia (32 x 4), muestra una asociación entre verracos y estirpe materna ($\chi^2_{93} = 420{,}5$). Esta asociación se confirma en el análisis separado de las tablas de contingencia en los verracos de cada estirpe con valores de χ^2 todos ellos significativos.

Métodos

Parámetros de cruzamiento

Para el análisis de este cruce dialélico completo se utilizó el modelo de EISEN y

Cuadro 1
Número de animales con registros de cada tipo de cruce entre las estirpes *Ervideira* (1), *Campanario* (2), *Caldeira* (3) y *Puebla* (4)
Table 1
Number of animals with records of each cross between the Ervideira *(1),* Campanario *(2),* Caldeira *(3) and* Puebla *(4) lines*

	Estirpe materna				Total
Estirpe paterna	1 (48)	2 (54)	3 (48)	4 (58)	(208)
1 (8)	64	31	31	30	156
2 (7)	29	70	18	28	145
3 (9)	26	24	71	29	150
4 (8)	18	18	25	67	128
Total (32)	137	143	145	154	579

Entre paréntesis: n.º de verracos y cerdas madres utilizados como reproductores.

col. (1983) que describe la media del cruce de las líneas i x i' mediante la ecuación:

$$\mu + \frac{1}{2}(l_i + l_{i'}) + m_{i'} + k(\bar{h} + h_i + h_{i'} + s_{ii'} + r_{ii'}) + e_{ii'}$$

siendo μ = media de las estirpes puras; l_i = efecto directo de la línea i; m_i = efecto materno de la línea i; $k = 0$ si $i = i'$ y $k = 1$ si $i \neq i'$; h = heterosis media; h_i = heterosis media de la línea i; $s_{ii'}$ = heterosis específica del cruce i x i'; $r_{ii'}$ = efecto reciproco debido a la diferencia entre el cruce i x i' y el cruce i' x i.

El modelo incluye las siguientes restricciones ($n = 4$):

$$\sum_{i=1}^{n} l_i = \sum_{i=1}^{n} m_i = \sum_{i=1}^{n} h_i = \sum_{i=1}^{n} r_{ii'} = \sum_{i'=1}^{n} r_{ii'} =$$

$$= \sum_{i=1}^{n} s_{ii'} = \sum_{i=1}^{n} s_{ii'} = 0; \quad s_{ii'} = s_{i'i}; \quad r_{ii'} = -r_{i'i}$$

En un dialélico completo 4 x 4 solo son estimables 16 de los 26 parámetros del modelo anterior siendo los otros 10 combinaciones lineales. En nuestro caso el vector p de parámetros estimables incluye los siguientes: (μ, l_2, l_3 l_4, m_2, m_3, m_4, $\bar{h}$, h_2, h_3, h_4, s_{24}, s_{34}, r_{12}, r_{14}, r_{24}). El modelo y las restricciones anteriores pueden expresarse en forma matricial mediante la matriz S (26 x 26) que se estructura de la forma siguiente:

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

siendo S_{11} = matriz de 16 x 16 que relaciona los cruces con los parámetros estimables; S_{12} = matriz 16 x 10 que relaciona los cruces con los parámetros no estimables; S_{21} = matriz de 10 x 16 que describe las restricciones de los parámetros estimables; S_{22} = matriz de 10 x 10 que describe las restricciones de los parámetros no estimables.

La reparametrización del modelo supone la obtención de una matriz S* (16 x 16) que relaciona cruces y parámetros estimables con las restricciones absorbidas. Se calcula mediante $S^* = S_{11} - \left(S_{12}S_{22}^{-}S_{21}\right)$. De acuerdo con el procedimiento de estimación directa

de parámetros de cruzamiento propuesto por KOMENDER y HOESCHELE (1989), los elementos de esta matriz se incluyen como covariables en los modelos de análisis de las observaciones individuales.

Modelos

Se emplearon los modelos siguientes, que comparten los mismos efectos fijos y covariables:

Modelo I. Modelo unicarácter de efectos fijos

$$y = X\,\beta + Q\,(S^*\,p) + e$$

siendo y = vector de observaciones (P_{14M} o GMD), X = matriz de incidencia de los efectos fijos en β (año y sexo), Q = matriz de incidencia que relaciona individuos y tipo de cruce, S* y p los descritos en el anterior apartado y e = vector de errores aleatorios con $E(e)= 0$ y $var(e)= I\sigma^2_e$.

Modelo IIa. Modelo animal bivariante

$$Y = X\,b + Q\,(S^*\,p) + Z\,u + e$$

donde Y = matriz de observaciones (P_{14M}, GMD), X, β, Q, S* y p ya descritos, Z = matriz de incidencia de u, u = efectos aleatorios de los animales con distribución N(0, $A\otimes\Sigma u$) siendo A y Σ_u las matrices de relaciones de parentesco y de (co)varianzas genéticas aditivas y e = matriz de los errores aleatorios con distribución N(0, $I\otimes\Sigma_e$), siendo Σ_e la matriz de (co)varianzas residuales.

Modelo IIb. Modelo animal bivariante con análisis bayesiano

En este enfoque se ha aplicado el procedimiento propuesto por VARONA y col. (1997)

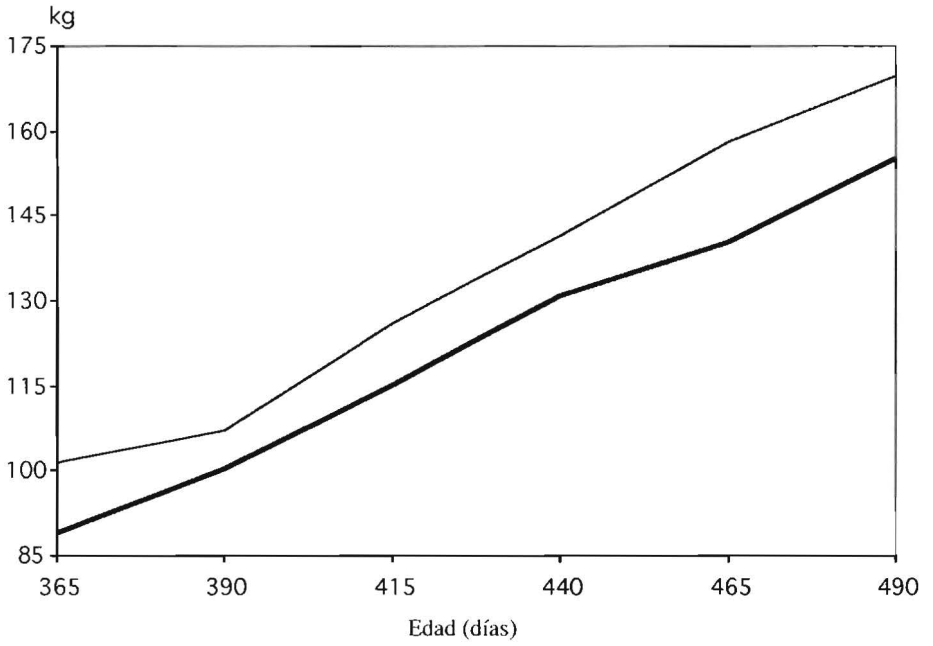


Figura 1. Peso medio en función de la edad de animales procedentes de líneas puras (—) y de cruces entre líneas (—)

Figure 1. Average weight by age of pure (—) and crossbred (—) pigs

que posibilita el ajuste de los parámetros de la función lineal de crecimiento (intercepto y pendiente) de forma conjunta a la estimación de los restantes parámetros. Intercepto y pendiente corresponden a los caracteres (P_{14M} y GMD) estimados mediante ajuste previo en los procedimientos anteriores. La elección de una función lineal para la descripción de crecimiento en montanera puede justificarse de acuerdo a los datos presentados en la figura 1. Se ha diferenciado la evolución del crecimiento de los animales puros y procedentes del cruce entre líneas para representar gráficamente el efecto de la heterosis media, uno de los parámetros que pretendemos estimar en este trabajo.

Modelo IIc. Modelo animal bivariante con análisis bayesiano y utilización de información *a priori*.

Mientras en el modelo IIb se asume una situación de máxima incertidumbre previa sobre los componentes de varianza de intercepto y pendiente, en este método se ha utilizado la información obtenida sobre (co)varianzas genéticas y residuales de los parámetros de la función lineal de crecimiento en montanera de cerdos ibéricos en un trabajo anterior (FERNÁNDEZ y col., 2000a). A partir de los coeficientes de variación de las correspondientes estimas, se calcularon los valores de los parámetros v_i que expresan el grado de incertidumbre de las mismas de acuerdo con el procedimiento propuesto por WANG y col. (1994). Entre los valores obtenidos para las (co)varianzas genéticas y residuales, se eligió en cada caso el menor que corresponde al mayor grado de incertidumbre.

Algoritmos

Para el análisis con los modelos I y IIa se utilizaron, respectivamente el procedimiento GLM de SAS y el programa VCE que estima componentes de varianza por maxima verosimilitud restringida y estima posteriormente los efectos fijos y covariables.

En el caso de los modelos IIb y IIc, las distribuciones marginales posteriores de los parámetros de interés se obtuvieron mediante el algoritmo de muestreo de Gibbs. Los detalles sobre este procedimiento pueden encontrarse en VARONA y col. (1997) y FERNÁNDEZ y col., (2000a). La longitud total de

Cuadro 2
Estimas de heredabilidad (h^2) y correlación genética (r_u) y residual (r_e) para el peso a los 14 meses (P_{14M}) y ganancia media diaria (GMD)
Table 2
Estimates of heritability (h^2) and genetic (r_u) and residual (r_e) correlation for the weight at 14 months (P_{14M}) and average daily gain (GMD)

	Modelo IIa		Modelo IIb		Modelo IIc	
Parámetro	Media	d.t.	Media	d.t.	Media	d.t.
h^2_{P14M}	0,199	0,059	0,166	0,075	0,105	0,031
h^2_{GMD}	0,066	0,045	0,206	0,110	0,247	0,037
r_u	0,354	0,348	0,619	0,241	0,242	0,113
r_e	0,396	0,042	0,844	0,033	0,864	0,018

muestreo fue en ambos casos de 1.500.000 muestras, descartándose las 500.000 primeras y guardando una de cada 100 iteraciones sucesivas. Se verificó la convergencia a la distribución diana duplicando el muestreo desde diferentes valores iniciales.

Resultados y discusión

Las estimas de heredabilidad de los caracteres estudiados y de la correlación genética y ambiental junto a sus correspondientes desviaciones típicas obtenidas por los tres modelos que incluyen la matriz de parentesco se presentan en el cuadro 2. En la comparación de estos resultados debe advertirse que se basan en información parcialmente distinta. El ajuste previo de los datos de crecimiento que requiere el modelo IIa obliga a prescindir de 71 animales que cuentan con un solo registro disponible, lo que no permite establecer los valores de P_{14M} y GMD en estos individuos. Sin embargo, el procedimiento bayesiano utilizado en el modelo IIb permite el análisis de toda la información registrada en los datos, lo que constituye una de sus ventajas. El modelo IIc utiliza adicionalmente la información previa disponible respecto a los componentes de varianza de estos caracteres procedente de registros de montanera de 762 cerdos ibéricos.

En la figura 2 se presentan las distribuciones a priori de las heredabilidades del intercepto y de la pendiente, con valores de la media y desviación típica posteriores de 0,492 (0,119) y 0,617 (0,094) respectivamente, medias muy superiores a las inferidas a partir de los datos (modelos IIa y IIb). Por el contrario, la correlación genética *a priori*

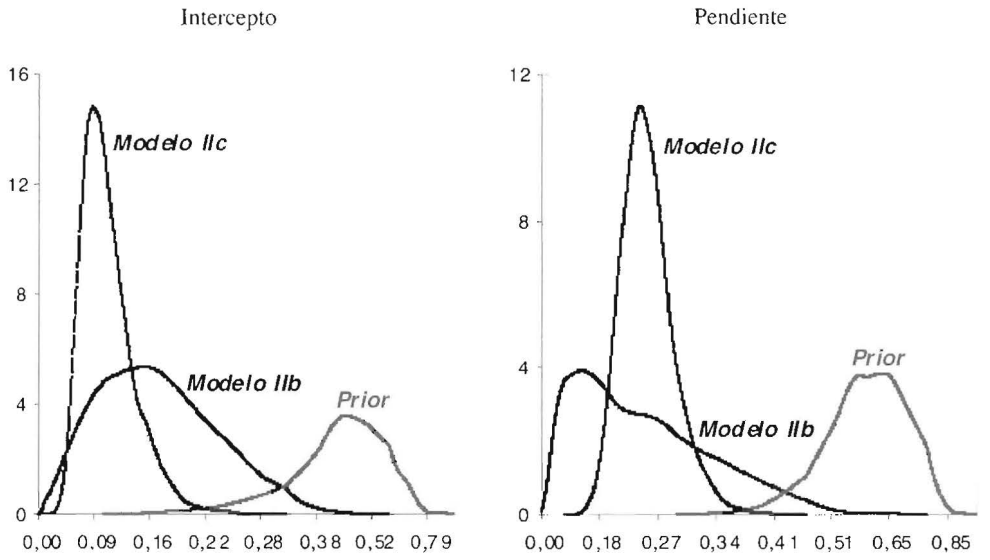


Figura 2. Distribución a *priori* y distribuciones marginales posteriores de la heredabilidad del intercepto y pendiente bajo los modelos IIb y IIc

Figure 2. A priori *distribution and marginal posterior distributions of heritability of intercept and slope under the models IIb and IIc*

entre ambos (distribución no representada) es superior a la inferida de los datos de este trabajo: 0,058 (0,172) (FERNÁNDEZ y col., 2000a). En dicha figura se representa adicionalmente la relación de las distribuciones a priori de h^2 con las distribuciones marginales posteriores de estos parámetros considerando además de los datos la citada información previa (modelo IIc). Para ambos caracteres se observó, como era de esperar, una drástica reducción de la dispersión. Existe asimismo un efecto sobre la media de las distribuciones de difícil interpretación al tratarse de un análisis bivariante.

Los resultados de los parámetros de cruzamiento estimados mediante el modelo genético propuesto por EISEN y col. (1983) se resumen en los cuadros 3 y 4 para el peso a los 14 meses y ganancia media diaria, respectivamente. En ellos se incluye junto a las estimas de la media de las estirpes puras y los efectos de líneas estimables, los efectos de aquellos parámetros significativamente diferentes de cero con alguno de los modelos analizados.

Como ya indicaron KOMENDER y HOESCHELE (1989), el modelo de efectos fijos subestima los errores de los parámetros estimados con respecto a los modelos que incluyen la matriz de parentesco. Del mismo modo el grado de incertidumbre de los componentes de varianza que asumen los análisis bayesianos también influye en las estimas de los errores siendo menores en el modelo IIc que en el IIb , y estos a su vez mayores a los obtenidos en IIa que supone el conocimiento de los verdaderos valores de los componentes de varianza.

La disparidad entre los diferentes modelos en los valores estimados de los parámetros de cruzamiento, tanto para P_{14M} como para GMD puede explicarse en parte por la diferente información utilizada en los modelos I y IIa y en los modelos IIb y IIc, así como por los diferentes valores de heredabilidad asumidos en cada uno de ellos, dado que la distribución de las progenies de los reproductores por tipo de cruzamiento no fue equilibrada (KOMENDER y HOESCHELE, 1989).

Cuadro 3
Estimas de los parámetros de cruzamiento más relevantes para el peso a los 14 meses (kg)
Table 3
Estimates of the most relevant crossbreeding parameters for weight at 14 months (kg)

	Modelo I	Modelo IIa	Modelo IIb	Modelo IIc
Parámetro	Media (d.t.)	Media (d.t.)	Media (d.t.)	Media (d.t.)
μ	126,95 (1,10)	123,88	121,13 (2,80)	120,88 (2,47)
l_2	-7,31 (2,67)	-7,29 (3,50)	-11,61 (5,47)	-11,83 (4,69)
l_3	-2,16 (2,67)	-2,32 (3,78)	1,35 (5,93)	2,11 (5,19)
l_4	1,80 (2,71)	0,19 (3,60)	1,51 (5,92)	1,82 (5,05)
$\bar{h}$	12,39 (1,42)	14,74 (1,85)	19,35 (3,05)	19,32 (2,90)
h_4	-3,55 (1,65)	-2,21 (1,56)	-5,11 (2,58)	-5,09 (2,42)
r_{24}	-2,63 (1,62)	-2,93 (1,54)	-5,15 (2,56)	-5,17 (2,43)

Los resultados obtenidos con el modelo de efectos fijos reflejan una aparente ventaja de la línea 4 (*Puebla*) para ambos caracteres, así como un posible efecto materno de la misma sobre el crecimiento diario. Este último resultado es difícil de explicar en términos biológicos y la supuesta superioridad de *Puebla* contradice estudios más amplios de comparación de estas mismas líneas (GARCÍA-CASCO, 1993). Ambos efectos no fueron detectados cuando se utiliza el modelo animal, especialmente con un enfoque bayesiano.

Cabe destacar el alto grado de significación estadística de la heterosis media en todos los casos y para los dos caracteres analizados. Este parámetro representa la diferencia entre la media de los individuos cruzados y la media de los individuos puros por lo que indicaría las diferencias en el crecimiento de ambos tipos de animales. El valor obtenido para crecimiento diario es equivalente al promedio descrito de heterosis entre razas de cerdos (GUNSET y ROBINSON, 1990). Si expresamos la heterosis como porcentaje de la media de las estirpes puras, los valores estimados con los diferentes modelos para el peso a los 14 meses de edad oscilan entre 9,8 y 16,0%, y para la ganancia media diaria entre 5,5 y 10,0%.

Con un procedimiento equivalente al modelo IIb, FERNÁNDEZ y col. (2000b) analizaron datos de crecimiento anterior a la ceba en animales procedentes del cruce dialélico de estas mismas estirpes, sometidos a dos planos de alimentación diferentes: alimentación fuertemente restringida y *ad libitum*. En los primeros la heterosis media fue 15,9% mientras en los segundos fue solo de 10,7%. Si asimilamos estos niveles de alimentación a un medio adverso en el primer caso y óptimo en el segundo, estos resultados junto a los obtenidos en este trabajo para el periodo final de ceba *ad libitum*, son congruentes con la hipótesis que relaciona heterosis y resistencia al medio adverso. Antecedente ilustre de esta relación es el libreto de *La forza del destino*, ópera de Verdi basada en una tragedia del duque de Rivas. Su protagonista, don Álvaro, es un mestizo hijo de hidalgo español e india americana, que supera con éxito a lo largo de los cuatro actos todo tipo de calamida-

Cuadro 4
Estimas de los parámetros de cruzamiento más relevantes para la ganancia media diaria (kg/d)
Table 4
Estimates of most relevant crossbreeding parameters for average daily gain (kg/d)

	Modelo I	Modelo IIa	Modelo IIb	Modelo IIc
Parámetro	Media (d.t.)	Media (d.t.)	Media (d.t.)	Media (d.t.)
μ	0,583 (0,007)	0,527	0,519 (0,016)	0,518 (0,017)
l_2	-0,009 (0,017)	0,009 (0,020)	-0,021 (0,031)	-0,013 (0,031)
l_3	-0,037 (0,018)	-0,054 (0,021)	-0,033 (0,034)	-0,035 (0,037)
l_4	0.038 (0,018)	0,031 (0,021)	0,041 (0,033)	0,042 (0,033)
m_4	-0.031 (0,013)	-0,021 (0,014)	-0,033 (0,019)	-0,034 (0,018)
$\bar{h}$	0,055 (0,009)	0,029 (0,013)	0,052 (0,016)	0,052 (0,016)
h_4	-0,024 (0,011)	-0,019 (0,010)	-0,032 (0,014)	-0,030 (0,014)

des: exilios, guerras, duelos con el bajo y el barítono y la muerte final de su amada Leonora. No parece, por tanto, exagerado reivindicar la expresión *fuerza del mestizo* en lugar del descafeinado vigor híbrido adecuado en todo caso a las mazorcas de maíz del Medio Oeste americano.

Agradecimientos

El presente trabajo se ha realizado en el marco del proyecto 1FD97-0772.

Bibliografía

COMERFORD J.W., BENYSHECK L.L., BERTRAND J.K. y JHONSON M.H., 1988. Evaluation of performance characteristics in a diallel among Simmental, Limousin, Polled Hereford and Brahman beef cattle. I. Growth, hip height and pelvic size. J. Anim. Sci. 66, 293-

EISEN E.J., HÖRSTGEN-SCHAWARK G., SAXTON A.M. y BANDY T.R., 1983. Genetic interpretation and analysis of diallelic crosses with animals. Theor. Appl. Genet. 65, 17-23.

FERNÁNDEZ A., SILIÓ L. y RODRÍGUEZ M.C., 2000a. Análisis genético del crecimiento en montanera de cerdos ibéricos. ITEA 96(1). 77-89.

FERNÁNDEZ A., RODRÍGUEZ M.C., RODRIGAÑEZ J., SILIÓ L. y TORO M.A., 2000b. Use of Bayesian *analysys* of growth functions to estimate crossbreeding parameters in Iberian pigs. 51th Annual Meeting of the European Association of Animal Production (EAAP). La Haya.

GARCÍA-CASCO J.M, 1993. Aspectos genéticos de la mejora de caracteres de crecimiento en cerdos Ibéricos. Tesis Doctoral. Universidad Complutense de Madrid.

GUNSET F.C. y ROBISON O.W., 1990. Crossbreeding effects on reproduction, growth and carcass traits. En Genetics of swine (Ed L.D. Young), USDA, Nebraska, 110-118.

KOMENDER P. y HOESCHELE I., 1989. Use of mixed-model methodology to improve estimation of cross-breeding parameters. Livest. Prod. Sci. 21, 101-113.

VARONA L., MORENO C., GARCÍA-CORTÉS L.A. y ALTARRIBA J., 1997. Multiple trait genetic analysis of underlying biological variables of production functions. Livest. Prod. Sci. 47, 201-209.

ITEA (2000), Vol. **96A N.° 3,** 270-276

ESTUDIO DEL CARÁCTER VELOCIDAD DE ORDEÑO Y RECUENTO DE CÉLULAS SOMÁTICAS EN PRIMERA LACTACIÓN DE LA POBLACIÓN VACUNA FRISONA DEL PAÍS VASCO

M.T. Fernández*, R. Alenda*, M. Serrano, E. Ugarte*****

* Departamento de Producción Animal, ETSIA, UPM, Madrid, España

** Departamento de Mejora Genética y Biotecnología, INIA, Madrid, España

*** NEIKER, S.A. Granja Modelo de Arkaute, Vitoria, España

RESUMEN

El objetivo de este estudio es conocer el valor genético de los caracteres velocidad de ordeño (V.O.) y recuento celular (RCS), de los toros utilizados como reproductores en la población frisona del País Vasco y la estima de los parámetros genéticos de estos caracteres. Se utilizó un total de 13.240 animales en primera lactación con datos de RCS y V.O. Con estos datos la estima de la heredabilidad mediante la metodología REML, fue de 0,10 y 0,17 para la V.O. y RCS respectivamente, siendo la correlación genética obtenida entre los caracteres de –0,41. Según los resultados sería conveniente realizar un Índice de sanidad de la ubre que contribuya a disminuir las pérdidas económicas ocasionadas por estos caracteres.

Palabras claves: Velocidad de ordeño, Recuento celular, Parámetros genéticos.

SUMMARY

The objetive of this study is the estimation of the breeding values of Freisian sires belong to Basque Country, for the traits milking speed and somatic cell count and genetic parameters for both traits. Data consist of 13.240 first lactation records of milking speed (MS) and somatic cell count (SCC). REML heritability estimates were 0,10 and 0,17 for MS and SCC respectively. Genetic corelation obtained was –0,41. This results indicate the advantaje to develop an Udder Health Index to reduce economic lost produced by these traits

Key words: Milking speed, Somatic cell counts, Genetic correlation.

Introduction

La velocidad de ordeño (V.O.) o tiempo de ordeño es un carácter de importancia económica en la producción de leche. Aumentar la velocidad de ordeño por vaca, conlleva una disminución del tiempo que se dedica a esta labor, lo cual supone un coste alto dentro de las explotaciones de vacuno lechero. Esto hace que sea un carácter cada

vez más considerado dentro de los programas de mejora.

Otro aspecto a destacar de la velocidad de ordeño como carácter a incluir en un índice de selección, es la correlación que tiene con el recuento de células somáticas. BOETTCHER y col. (1998), encontraron una correlación genética entre velocidad de ordeño y recuento celular en primera lactación de –0,41, lo que significa que a mayor velocidad de ordeño, mayor es el RCS. En relación con el carácter V.O., hay que decidir que interesa más, si seleccionar animales con alta V.O. o con baja V.O. Si la V.O. es alta, disminuye la tensión del esfínter del pezón, lo que provoca una menor resistencia de la ubre a los patógenos y por tanto, un aumento del riesgo de mamitis clínicas (BOETTCHER y col., 1998). Por el contrario, un animal con V.O. excesivamente lenta con respecto a sus compañeras de sala de ordeño, significa que en este animal no termina bien la estracción completa de la leche lo cual provoca un aumento de tensión intramamaria y los consecuentes problemas de secado y posterior aparición de la infección. Por tanto parece razonable considerar la velocidad de ordeño como un carácter de óptimo intermedio.

Aspectos relacionados directamente con la V.O., son la propia máquina de ordeño y la rutina de ordeño, y el propio animal. Del primer grupo hay que destacar la influencia de las técnicas de ordeño y el nivel de vacío de la máquina. Trabajar con un nivel de vacío superior al especificado supone un aumento en el flujo de la leche, pero trae consigo un agotamiento de las vacas por irritar el recubrimiento del pezón. En cuanto los factores relacionados con el propio animal, son importantes el nivel de producción y la sanidad de la ubre (DUDA, 1996). Las vacas de elevada producción presentan una mayor presión intramamaria, que afecta de modo directo al grado de apertura del esfínter y nivel de secreción hormonal y de forma global al estado general de la vaca. En cuanto a la sanidad de la ubre, existe una estrecha relación con la velocidad de ordeño. Las vacas con pezones de canales amplios, requieren menor manipulación en el ordeño, por tanto tienen mayor velocidad. Por el contrario, esta mayor apertura del esfínter del pezón, trae consigo una mayor facilidad para la entrada de patógenos, lo que supone un mayor riesgo de mamitis (BOETTCHER y col., 1998).

Por estos motivos, se ve clara la necesidad de definir un índice de Sanidad de la Ubre, que tenga en cuenta los caracteres mencionados, V.O y RCS, y que considere también el carácter de conformación de ubres debido a la gran correlación que existe entre estos tres caracteres. BOETTCHER y col. (1997) realizaron un índice de sanidad, considerando caracteres de conformación de la ubre, V.O. y RCS. En este estudio se encontró una correlación entre V.O. y RCS en primera lactación de 0,43 mientras que la correlación entre RCS y los caracteres de conformación de la ubre oscilaba entre –0,20 y 0,26. En 1997, GRENGLER y GROEN (1997) analizaron las ventajas de seleccionar animales resistentes a mamitis utilizando un modelo multicarácter que incluía algunos caracteres relacionados con la conformación de las ubres y el RCS, encontrando una correlación de –0,70 entre resistencia a mamitis y RCS. La correlación encontrada para los caracteres de conformación de ubres con RCS oscilaba entre 0,30 a 0,40.

El índice de sanidad de la ubre contribuye a disminuir los costes ocasionados en las explotaciones por la incidencia de mamitis y por tanto a aumentar la rentabilidad de las mismas. En España, se estima que un RCS de 500.000 céls/ml. equivale a una pérdida aproximada de 3 ptas/litro (TENA y col., 1999). Según esto, sería interesante seleccionar animales cada vez más resistentes a la

enfermedad y con una velocidad de ordeño media de acuerdo con las condiciones de producción. SIVARAJASINGAM y col. (1984) y MEYER y col. (1987) ya daban a la V.O. el tercer puesto en importancia después de caracteres como producción de leche y contenido de grasa.

La V.O es por tanto un carácter que interesa bastante a los ganaderos. Desde finales de los 80 y principios de los 90, cada vez son más los países que lo consideran dentro de sus índices de selección. En la UE países como Francia, Italia, Holanda, Gran Bretaña, Alemania; y países fuera del ámbito comunitario como la Rep. Checa, Australia y Canadá, incluyen el carácter V.O en sus valoraciones.

En este estudio, se han estimado los parámetros genéticos para los caracteres de RCS y V.O. correspondientes a los datos en primera lactación de la raza Frisona utilizada en el País Vasco.

Material y métodos

Recogida de información

El objetivo de este estudio es conocer el valor genético que tienen los toros utilizados como reproductores en la población frisona del País Vasco para el carácter V.O.y RCS.

Para el carácter V.O. se han utilizado un total de 14.151 datos correspondientes a vacas de primera lactación, recogidas entre los años 1989 y 1999, y distribuidas en un total de 798 rebaños. La genealogía la constituyeron 33.864 animales. El carácter RCS, fue analizado en un total de 224.370 lactaciones. Dichos datos provienen de los controles que se realizan de cada lactación, desde el año 1985 hasta 1999, obteniéndose posteriormente una media de recuento transformado por lactación. En este caso se consideraron 1.336 rebaños. En este caso el registro genealógico consistió en 119.292 animales.

Para realizar la estima de los parámetros genéticos, se escogieron los animales que tuvieran dato de RCS y de V.O. en primera lactación, esto supuso un total de 13.240 registros. El efecto fijo rebaño-año tenía 2.862 niveles. En la genealogía se consideraron un total de 33.864 animales.

Definición del carácter

En España, para realizar la primera valoración de V.O., se ha utilizado una escala subjetiva tomada por el ganadero. Dicha escala puntúa al animal de 1, ordeño fácil, 2 ordeño medio, y 3 difícil o lento. Este dato se toma durante el 2.° o 3.° control correspondiente a la primera lactación. El hecho que este dato sea tomado por el ganadero puede suponer una ventaja con respecto a una sola medida objetiva, tomada por un controlador, ya que el ganadero evalúa el comportamiento de la vaca a lo largo de varios días de ordeño.

Al tratarse de valores discontinuos, lo habitual es realizar una transformación de Snell, (BOETTCHER *et al.*, 1998 y MEYER, 1987) que permita trasformar los datos en una escala de frecuencias normal. Este dato transformado es el que se utiliza en la valoración genética del carácter. Sin embargo, cuando se trata de transformar una escala categórica de sólo tres valores, la distribución de frecuencias según Snell, no interesa, puesto que no se obtiene un discernimiento aclaratorio con respecto a los valores sin transformar.

En cuanto al dato de recuento celular, es necesario transformarlo a una escala logarítmica debido al gran rango de variación que presenta. La transformación más frecuente es RCS = $\log_2$ (scc/100.000) + 3, (siendo scc valor correspondiente de recuento celular de cada control y RCS el valor transformado). Con esto se consigue una distribución parecida a la normal y el test de heterogeneidad de varianza (X^2) no es significativo (ALI y SHOOK, 1980). Una vez realizada la transformación, se calcula una media por lactación.

Metodología

Valoración genética

La valoración genética de los caracteres de V.O. y de RCS se realizó mediante la metodología BLUP modelo animal. Los programas informáticos utilizados para resolver las ecuaciones del modelo animal para ambos caracteres, fue la cadena de programas BLUP AM, (JURADO y col., 1992). Estas ecuaciones, se resuelven mediante el método iterativo de GAUSS-SEIDEL, con un total de 10.000 iteraciones. Esta cadena de programas, obtiene la fiabilidad de las valoraciones aplicando el método aproximado de MEYER. La heredabilidad considerada para ambos caracterse fue de 0,15 (MEYER, 1987).

Para el carácter V.O., se utilizó un modelo animal que incluyó como efectos fijos el rebaño-año y la estación de parto y como efecto aleatorio el genotipo del animal. La base de datos se restringió a aquellos grupos de comparación con un mínimo de 5 observaciones. Las estimas de la estación de parto, dada la imposibilidad de considerarlas mes a mes por la escasez de observaciones en cada nivel, se distribuyó en 3 clases, de marzo a junio, de julio a octubre, y de noviembre a febrero, en función de los resultados obtenidos de los análisis GLM (SAS procedure, 1995). Se estudió también la inclusión en el modelo de la producción de leche como covariable, sin embargo, se comprobó, con GLM (SAS procedure, 1995) que dicho efecto no era significativo, como cabría esperar puesto que se trata de una medida del carácter totalmente subjetiva recogida durante las observaciones correspondientes de dos ordeños diarios, y que no es función de la cantidad de leche producida durante la lectura del carácter. Según BOETTCHER y col. (1998), la producción de leche, tiene un efecto muy pequeño sobre la velocidad de ordeño, la V.O. aumenta poco con el aumento de producción.

La ecuación del modelo utilizado fue la siguiente:

$$Y_{ijkm}=\mu + RA_j+E_k+G_m+e_{ijkm}$$

Siendo: y: vector de observaciones, corresponde al dato de velocidad de ordeño (i=1,.....14.151); μ: media poblacional; RAj: efecto rebaño-año (8.778 niveles, correspondientes a 798 rebaños y 11 años); E_k: estación de parto (3 niveles); Gm:genotipo del animal; eijkm: error.

Para valorar el carácter de RCS, se utilizó el modelo animal con medidas repetidas que incluía como efectos fijos Rebaño-año, número de lactación y como efectos aleatorios el genotipo del animal y el efecto ambiental permanente. El dato utilizado para la valoración de RCS, corresponde a una media aritmética por lactación. Según esto, y como ocurre en otras especies (CHARFFEDINE, 1998; BANOS y SHOOK, 1990; ZHANG y col., 1994) es necesario hacer una precorreción del modelo según las estimas calculadas utilizando el procedimiento GLM (SAS procedure, 1995) de los efectos número de lactación-estado de lactación, fecha control, y como covariable la producción de leche por control.

La ecuación del modelo utilizado fue la siguiente:

$$Y_{ijkm}=\mu + RA_j+Nl_k+G_m+P_m+e_{ijkm}$$

Siendo: Y_{ijkm}: valor correspondiente a la media por lactación de la transformación logarítmica del número de células somáticas; μ: media poblacional; RA_j: efecto rebaño año (18704 niveles correspondientes a 1336 rebaños y 14 años); Nl_k: número de lactación (1:1ª lactación, 2:2ª lactación y 3: 3.ª y sucesivas lactaciones); G_m: genotipo del animal; P_m: efecto ambiental permanente; e: error.

Al igual que en el caso de los grupos de comparación en V.O., se exigió que el grupo rebaño-año tuviera como mínimo 5 observaciones por nivel.

Parámetros genéticos

La estima de los parámetros genéticos se llevó a cabo con modelos animales unicarácter y multicarácter para V.O y RCS, mediante metodología REML, utilizando el programa VCE (Groeneveld y Cortés, 1998). Para los análisis unicarácter se utilizaron los mismos modelos que en la valoración genética. Para el análisis multicarácter se utilizaron distintos modelos para V.O. y RCS, e iguales a los utilizados en la valoración genética, con la única excepción de que en el modelo para RCS no se incluyó el efecto número de lactación, ya que todos los regsitros eran de primera lactación.

Resultados y discusión

Valoración genética

Al estudiar la correlación existente entre 105 machos valorados para ambos caracteres en primera lactación (SAS procedure, 1995), se ha obtenido una correlación de –0,27. Este valor bajo y negativo de la correlación indica que las ordenaciones de los machos para los valores genéticos de V.O. y RCS son muy distintas y de sentido inverso. Como se trata de caracteres relacionados de forma negativa, implica, que menor valoración del carácter velocidad (más facilidad de ordeño), mayor será el valor de recuento celular y por tanto mayor el riesgo de padecer mamitis. Otros estudios de Zhang y col. (1994), obtuvieron un coeficiente de correlación desfavorable de 0,20 entre ambos caracteres para primera lactación.

Parámetros genéticos

Una vez realizada la valoración genética para los caracteres anteriores, se realizó una estima de los parámetros genéticos. Son numerosos los trabajos que han analizado V.O. y RCS, con el objetivo de buscar caracteres que de una forma indirecta aumenten la resistencia a la mamitis. Nielsen y col. (1996), encontraron una correlación genética entre RCS e incidencia a mamitis, significativa (0,30 a 0,97), esto implica que una selección a favor de bajo RCS, mejora la resistencia a mamitis. Shook (1989), Rogers y col. (1989) y Emanuelson (1997) encontraron una correlación genética importante y negativa entre SCS (somatic cell score) y la resistencia a mamitis. La escala utilizada en el SCS varía de 0 a 9, cada punto de variación en la escala, supone la duplicación de células somáticas.

En cuanto a la V.O., es importante destacar, que una elevada velocidad de ordeño, supone una menor resistencia de la ubre a los patógenos (Boettcher y col., 1998). Sin embargo Rupp and Boichard (1999), no encontraron una correspondencia clara entre

Cuadro 1
Heredabilidad, h^2, (diagonal) y correlación genética, r_g, (sobre la diagonal) para los caracteres velocidad de ordeño (V.O.) y recuento de células somáticas (RCS) estimadas mediante modelos unicarácter (UC) y multicarácter (MC)

Carácter	V.O. (MC)	RCS (MC)	V.O.(UC)	RCS (UC)
V.O.	0,11 ± 0,013	-0,41 ± 0,014	0,11 ± 0,014	
RCS		0,17 ± 0,019		0,18 ± 0,021

la V.O., y la existencia de mamitis clínica, aunque animales con V.O. alta, si presentaban un valor de RCS transformado suficientemente elevado.

En el cuadro 1, se presentan los parámetros genéticos obtenidos al analizar los caracteres de V.O. y RCS en primera lactación, utilizando modelos unicarácter y multicarácter.

La heredabilidad obtenida para RCS es similar a la obtenida por RUPP y col. (1999) (h^2 = 0,17) y algo inferior para V.O. (h^2 = 0,17), y superior a las encontradas en otras referencias para RCS, que oscilan entorno a 0.11 (SCHUTZ y col. 1990, ROGERS y col., 1991, DA *et al.*, 1992).

Conclusiones

Según DENTINE y col., (1987), un tercio de los animales desechados es debido a problemas relacionados con mamitis. La mamitis supone una pérdida de aproximadamente un 25% de los ingresos obtenidos por venta de leche, teniendo en cuenta los costes de medicamentos, veterinario, pérdidas por venta de leche, reducción por precio del litro de leche y desecho anticipado (BÉJAR y ALENDA, 1996).

Ala vista de los resultados, seleccionar animales con una baja V.O, supone tener animales con menor RCS, sin embargo, esto trae consigo un aumento de los costes de la rutina de ordeño. Por tanto, el óptimo de ese carácter parece ser una velocidad intermedia.

La heredabilidad para estos caracteres, 0,10 para V.O. y 0,17 para RCS, así como la correlación genética que presentan ambos caracteres pone de manifiesto la necesidad de elaborar un índice de sanidad de la ubre que contribuya a disminuir los costes ocasionados por mastitis dentro de una explotación. Además existen numerosas referencias (BOETTECHER y col.,1998 y RUPP y BOICHARD, 1999) en que se consideran dentro de este índice los caracteres de conformación de ubre, ya que también presentan una alta correlación con los caracteres antes mencionados.

Bibliografía

ALI A.K.A., SHOOK G.E., 1980. An optimun transformation of somatic cell concentration in milk production. J. Dairy Sci. 63:487.

BANOS G., SHOOK G.E., 1990. Genotype by environment interaction and genetic correlations among parities for somatic cell count and milk yield. J. Dairy Sci. 73:2563 .

BOETTCHER P.J, DEKKERS J.M.C., KOLSTAD B.W., 1998. Development of an Udder health index for sire selection based on somatic cell score, udder conformation, and milking speed. J. Dairy Sci. 81:1157.

BOETTCHER P.J., DEKKERS J.C.M., KOLSTAD B.W., 1997. An Udder Health Index for Sire Selection Based on Somatic Cell Score, Udder Conformation, Milking speed. Interbull n.º 15.

CHARFEDDINE N., 1998. Selección por Mérito Económico Global en el ganado vacuno frisón en España. Tesis Doctoral. E.T.S.I.A. UPM. Madrid.

DA I., GROSSMAN, M., MISZTAL I., WIGGANS G.R., 1992. Estiamtion of genetics parameters fro somatic cell score in Holstein. J. Diary Sci., 75: 2265.

GENGLER N., GROEN F., 1997. Potential Benefits from Multitrait Evaluation – an Example in Selection for mastitis Resistance Based on Somatic Cell Score And Udder Conformation.A Simulation study. Interbull n.º 15.

DUDA J., 1996. New prospects in sire evaluation for milkability.International Woerkshop of Genetic Improvement of Functional Traits in Cattle.Faculte Universitiare des Science Agronomiques, Gemboux, Belgium, January 21-23.

JURADO GARCÍA J.J, HERNÁNDEZ D., SERRANO M., 1992. 2.º Catálogo de software de interés en agricultura. Mayo 1992. Página 142, programa 248 BLUPAM.

MEYER K., BURNSIDE E.B., 1987. Scope for a subjective assesment of milking speed. J. Dairy Sci. 70: 1061.

ROGERS G.W., HARGROVE G.L., LAWOR T.L., EBERSOLE J.L., 1991. Correlations among linear types traits and somatic cell score. J. Diary Sci.74: 1087.

RUPP R., BOICHARD D., 1999. Genetic Parameters for Clinical Mastitis, Somatic Cell Score, Production, Udder Type Traits, and Milking Ease in First Lactation Holstein. J. Dairy Sci. 82: 2198.

SAS/STAT. User's Guide, Relase 6.11. 1995. SAS Inst., Inc., Cary, NC.

SCHUTZ M.M., HANSEN L.B., STEUERNAGEL G.R., RENAU, LUCK A.L., 1990. Genetics parameters for somatic cells, Protrein, and fat in milk on Holsteins.J. Dairy Sci. 73: 494.

SHOOK G.E., 1989. Selection for disease resistance. J. Dairy Sci. 72: 1349.

SIVARAJASINGAM S., BURNSIDE E.B., WILTO J.V., PFEIFFER W.C., GRIEVE D.G., 1984. Ranking dairy sires by linear programming dairy farm models. J. Dairy Sci. 67: 3015.

SNELL E.J., 1964. A scalling procedure for ordered categorical data. Biometrics 20: 592

TENA FERNÁNDEZ S.M., DE TENA ANDREU S., GASCA ARROYO A., 1999. Recuento de células somáticas y mamitis bovinas. Frisona Española, n.º 112.

GROENEVELD E., GARCÍA CORTÉS A., 1998. VCE 4.0. A covarinace component package for frecuentist and Bayesians. 6th World Congress on Genetics Applied to Livestock Production. Armidale, Australia.

ZHANG W.C., DEKKERS J.C.M., BANOS G., BURNSIDE E.B., 1994. Adjustment factors and genetic evaluation for somatic cell score and relationship with other traits of Canadian Holstein. J. Dairy Sci. 77: 659.

ITEA (2000), Vol. **96A N.° 3,** 277-281

IDENTIFICACIÓN DE ALELOS CON EFECTO SIGNIFICATIVO EN LA DIFERENCIACIÓN DE POBLACIONES

D. García*, C. Carleos, J.A. Baro***, J. Cañón***

* Departamento de Producción Animal, Facultad de Veterinaria, Avda. Puerta de Hierro s/n, 28040 Madrid, España

** Departamento de Estadística, Universidad de Oviedo, España

*** Unidad de Mejora Genética, SERIDA, España

RESUMEN

Los sistemas de clasificación racial basados en frecuencias de los alelos de marcadores y genotipos individuales se emplean en programas de conservación, detección de fraudes, análisis de poblaciones híbridas, estudios forenses y pruebas de paternidad. La técnica –asignación de individuos a una raza maximizando la probabilidad de que el individuo pertenezca a ella– es simple, aunque su aplicación rigurosa supone cierta complejidad. Este trabajo revisa uno de los aspectos críticos: la presencia de alelos atípicos. Éstos se definen como los que tienen frecuencias extremas en alguna raza; casos particulares son los alelos específicos de raza y los alelos ausentes en una raza. El enfoque que se ha seguido consistió en examinar para cada locus el comportamiento de una tabla de contingencia con las razas frente al número de alelos. La significación de cada tabla y de cada celda de cada tabla se estimó mediante tres métodos distintos: el habitual, basado en el estadístico Chi-cuadrado, y dos aproximaciones por simulación de contrastes exactos: MC y MCMC. En los casos simulados, la significación de la tabla de contingencia se estima como la proporción de simulaciones en que la probabilidad de la tabla observada es superior a la de la simulada. Se estudia la posibilidad de emplear un enfoque análogo para determinar una significación por celda.

Palabras clave: Marcadores, Asignación racial, Significación.

ABSTRACT

Breed assignment procedures based on marker allele frequencies and individual genotypes can be used for conservation purposes, fraud detection, analysis of hybrid populations, forensic advice, and paternity tests. The technique –allocation of an individual to the breed maximising the probability of that individual belonging to it– is straightforward but rigorously speaking it has a number of weak points. We address one of them in this paper: the presence of outlier alleles. Outlier alleles are those having extreme frequencies in a certain breed, with breed specific alleles and missing alleles as particular cases. The approach followed to study this topic was to examine for each locus the behaviour of breed vs. number of alleles contingency tables. The significance of each table and of each cell in each table was estimated by three different methods –the usual chi-square statistic and two simulated –MC and MCMC– exact tests. In the simulated cases significance was estimated as the proportion of realisations where the

probability of the observed table exceeded that of the simulated one, proceeding analogously for individual cells.

Key words: Markers, Breed allocation, Significance.

Introducción

Por asignación racial se entiende la clasificación de una muestra de un animal en alguna de una serie de razas a las que el animal pueda o haya podido pertenecer. En los últimos tiempos se han venido desarrollando diferentes técnicas de asignación que tienen en cuenta el genotipo de la muestra y las frecuencias de los alelos de marcadores en las diferentes razas (BANKS & EICHERT 2000, SHRIVER *et al.* 1997). Algunas de las aplicaciones de estas técnicas son: asignar individuos a razas en programas de conservación étnica, detección de fraudes, análisis de las proporciones de las poblaciones originales en las híbridas, monitorización de la inmigración, estudios forenses y pruebas de paternidad. La información genética empleada (ver la comparación realizada por BLOTT, 1999) puede ser de marcadores multialélicos como microsatélites, de alta disponibilidad, con métodos estándar de genotipado y muy informativos, o basada en marcadores dialélicos como los SNPs, que son mucho más frecuentes y de los que se espera que pronto permitan realizar genotipados masivos a bajo coste.

Métodos

La información sobre microsatélites y otros polimorfismos similares se analiza con métodos estadísticos clásicos de teoría de la decisión para resolver el problema de la asignación racial. El criterio de asignación consiste en asignar un individuo a la raza en la que la aparición de ese individuo sea más probable. Ésta se toma como la que maximiza la probabilidad de aparición de ese individuo, dado un conjunto de razas y de loci y conocidas las frecuencias de los alelos de dichos loci en cada raza. Se asumirán equilibrios de ligamiento y de Hardy-Weinberg. Nos centraremos en el estudio de la presencia de alelos especialmente influyentes en la clasificación: los alelos atípicos que presentan frecuencias particularmente altas o bajas en alguna población. Casos extremos los constituyen los alelos que aparecen asociados únicamente a una raza, y los alelos que no han sido registrados en alguna raza. La hipótesis nula de partida es la homogeneidad de poblaciones respecto a la distribución de los alelos. Los datos del análisis se presentan en una tabla de contingencia "poblaciones / alelos". Como ejemplo, la siguiente tabla recoge las frecuencias alélicas observadas para el locus CSSM en 50 individuos de las razas bovinas 'asturiana de los valles', 'asturiana de la montaña' y 'pirenaica':

ASM	0	3	0	10	13	5	31	2	10	2	18	6
ASV	1	1	3	17	10	4	24	6	16	1	8	9
PIR	0	6	1	14	40	0	16	0	10	3	4	6

Existen varios enfoques para analizar la significación de las celdas (casillas) bajo la hipótesis de homogeneidad. En primer lugar, comentaremos tres estrategias de contraste empleadas con la tabla en conjunto. El problema radica en contrastar la hipótesis de distribución homogénea de los alelos en todas las poblaciones, lo que equivale a con-

trastar las hipótesis de independencia entre filas y columnas. El estadístico tradicionalmente empleado es el Chi-cuadrado:

$$X = \sum_{Celdas} \frac{(obs. - esp.)^2}{esp.}$$

Como regla general, no se recomienda su uso cuando hay celdas con menos de 5 observaciones. En tal caso, en una tabla 2 x 2 es factible emplear el método exacto de Fisher, pero con tablas mayores, como es el caso de los microsatélites, que son muy polimórficos, el número de combinaciones lo hace computacionalmente impracticable. Se ha de recurrir a procedimientos de Montecarlo. Se describirán dos de ellos, que denominaremos MC (Montecarlo clásico) y MCMC (Montecarlo con cadenas de Markov).

Montecarlo clásico. El algoritmo para el cálculo de la significación de una tabla frente a la hipótesis nula de independencia u homogeneidad es como sigue:

1. Calcular la probabilidad de la tabla observada T :

$$Pr[T] = \frac{\prod_{i=1}^{nfilas}(T_{i\cdot}!)\prod_{j=1}^{ncols}(T_{\cdot j}!)}{T_{\cdot\cdot}!\prod_{i=1}^{nfilas}\prod_{j=1}^{ncols}(T_{ij}!)}$$

2. Generar aleatoriamente N tablas bajo la hipótesis de independencia y con los mismos recuentos marginales que T: $T_1,...,T_N$.

3. Calcular las probabilidades de estas tablas.

4. Compararlas con Pr[T] y estimar la significación (p-valor) de T como:

$$\hat{P}_v = \frac{\#\{T_x \mid P[T_x] \le P[T], x \in \{1,\ldots,N\}\}}{N}$$

MCMC. Cuando se emplea simulación de Montecarlo con cadenas de Markov, la generación de nuevas tablas bajo la hipótesis nula es mucho más rápida. Con la tabla del ejemplo, la generación de 100.000 tablas mediante Montecarlo clásico llevó 3 min 52 s., en tanto que 10.000.000 tablas MCMC consumieron solamente 49 s. El algoritmo propuesto por Guo y Thompson (1992) cambia un alelo en cada iteración y por tanto sólo se modifican los valores de cuatro celdas respecto a la tabla generada en la iteración previa. La tabla de partida es la observada. Dados $1 \le i,i' \le nfilas$ y $1 \le j,j' \le ncols$, el cambio afecta únicamente a las casillas (i,j), (i',j), (i,j') y (i',j'). Aunque es mucho más rápido que el MC, se requiere la generación de muchas más tablas para obtener la misma precisión en las estimaciones, al no ser éstas independientes. Hastings (1970) demostró que en estas circunstancias puede obtenerse un estimador del p-valor, $\hat{P}_v$, asintóticamente normal y que converge en probabilidad a su valor real.

De la tabla a la celda

Se plantean ahora estrategias centradas en la significación de las celdas y no de las tablas. Se pretende determinar cuáles de las casillas de cierta tabla de contingencia son las responsables de que la tabla se aleje de la hipótesis nula de independencia.

Estadístico χ^2. La propia forma del estadístico sugiere asignar como "índice de rareza" o "atipicidad" el sumando correspondiente a cada celda. El estadístico χ^2 es la suma, celda a celda, de la diferencia cuadrática "tipificada" entre el valor observado de la celda y el valor esperado (el que tendría la celda, por término medio, bajo la hipótesis nula). Este valor ofrece una medi-

da comparativa de la contribución de cada celda al alejamiento de la hipótesis nula.

Métodos de Montecarlo. Por cada tabla generada en los métodos MC y MCMC, se registran los valores observados (k) en cada celda (i,j) y se incrementan unos contadores que acumulan el número de veces que la celda (i,j) adopta el valor k. Así se obtiene una estimación de la distribución de probabilidades de los diferentes valores que puede adoptar cada celda bajo la hipótesis nula. La distribución de los valores de una celda ofrece toda la información estadística sobre su significación. Para resumir la relación entre la distribución y el valor observado, existen varias opciones:

1. "Tipificación": si E es la media muestral, y D la desviación típica muestral, dar el valor

$$\frac{k - E}{D}$$

o su cuadrado como índice de atipicidad; este valor indica la rareza de los valores observados en la celda: cuanto mayores son, más atípico es el valor observado, por lo que permite la comparación entre celdas. Sin embargo, no tiene una significación propia nítidamente definida, salvo que se suponga que la distribución por celda es más o menos normal (una corrección por continuidad sería adecuada en este caso).

2. Estimar el "p-valor"; según la "definición" empleada, puede ser:

a) Calcular la suma de las probabilidades estimadas más bajas que la probabilidad estimada del valor k:

$$\sum_{\Pr[l]\le\Pr[k]} \Pr[l].$$

b) Calcular el doble del mínimo de: la suma de las probabilidades estimadas para los valores menores o igules que k, y la suma de las probabilidades estimadas para los valores mayores o iguales que k (MANLY *et al.* 1998):

$$2 \times \min\left\{\sum_{l\le k} \Pr[l], \sum_{l\ge k} \Pr[l]\right\}.$$

Los valores calculados en 2 pretenden estimar directamente una significación, pero como estimadores resultan inconsistentes en el caso de que exista al menos un valor con la misma probabilidad teórica que el observado: Sea k el valor observado en cierta celda. Sea P_v el p-valor asociado y $\hat{P}_v$ el estimador según 2a. Sean k y k' dos valores que puede adoptar la celda (i,j), y supóngase que $\Pr[k|H_0] = \Pr[k'|H_0]$. Sea p_k la estimación empírica de $\Pr[k|H_0]$ y $p_{k'}$ la de $\Pr[k'|H_0]$. Cuando el número de iteraciones es muy grande, $\Pr[p_k > p_{k'}] \approx 0'5 \approx \Pr[p_k < p_{k'}]$, con lo que existe $\varepsilon = p_k/2 > 0$ tal que $\Pr[|\hat{P}_v - P_v| > \varepsilon] \approx 0'5$ no tiende a cero y $\hat{P}_v$ no es débilmente consistente.

Resultados

A partir de la tabla de ejemplo, obtendremos cada uno de los estadísticos antes descritos. En cuanto a la significación global de la tabla frente a la hipótesis de homogeneidad, el estadístico χ^2 toma un valor de 67'79 (P_v=1'47E-6). Las estimaciones del p-valor mediante Montecarlo son las dos prácticamente nulas (MC: 1'0E-5, MCMC: 2'0E-7).

Los valores del estadístico χ^2 celda a celda fueron:

0'33	0'03	1'33	0'98	3'04	1'33	2'27	0'17	0'33	0'00	6'40	0'14
1'33	1'63	2'08	0'81	5'76	0'33	0'00	4'17	1'33	0'50	0'40	0'57
0'33	2'13	0'08	0'01	17'19	3'00	2'48	2'67	0'33	0'50	3'60	0'14

Los valores de Montecarlo (MC y MCMC) tipificados fueron:

0'50	0'05	2'02	1'69	5'71	2'04	4'48	0'26	0'57	0'00	10'68	0'24
2'01	2'54	3'14	1'40	11'23	0'52	0'01	6'42	2'27	0'76	0'65	0'93
0'50	3'29	0'12	0'01	32'52	4'62	4'84	4'09	0'57	0'76	6'01	0'23

Los p-valores estimados, fueron:

1'00	1'00	0'31	0'21	0'02	0'17	0'04	0'72	0'57	1'00	0'00	0'81
0'33	0'17	0'11	0'29	0'00	0'49	1'00	0'02	0'14	0'67	0'54	0'34
1'00	0'89	1'00	0'86	0'00	0'03	0'03	0'06	0'57	0'40	0'01	0'81

Conclusiones

Hemos querido obtener medidas objetivas para la identificación de las celdas "causantes" del alejamiento de la hipótesis de homogeneidad en tablas de contingencia. El estimador del nivel crítico (p-valor) propuesto no es consistente cuando existe un valor de la celda que bajo la hipótesis nula tiene la misma probabilidad que el observado. Aunque este hecho es presumiblemente infrecuente, conviene aportar adicionalmente algún valor de diagnóstico más estable, como los "índices tipificados".

Bibliografía

BANKS, M.A., EICHERT, W., 2000. WHICHRUN (version 3.2): A computer program for population assignment of individuals based on multilocus genotype data. Journal of Heredity, 91:87-89.

BLOTT, S.C., WILLIAMS, J.L., HALEY, C.S., 1999. Discriminating among cattle breeds using genetic markers. Heredity, 82:613-619.

GUO, S.W., THOMPSON, E.A., 1992. Performing the exact test of Hardy-Weinberg proportions for multiple alleles. Biometrics, 48:361-372.

HASTINGS, W. K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57: 97-109.

MANLY, B.F.J., 1998. Randomization, bootstrap and Monte Carlo methods on biology. Chapman & Hall.

SHRIVER, M.D., SMITH, M.W., JIN, L., MARCINI, A., AKEY, J.M., DEKA, R., FERRELL, R.E., 1997. Ethnic-affiliation estimation by use of population-specific DNA markers. American Journal of Human Genetics, 60:957-964.

ITEA (2000), Vol. **96A N.° 3,** 282-289

MEDIDAS DE DIVERSIDAD GENÉTICA EN POBLACIONES DE CABALLOS CELTAS ESPAÑOLES

D. García, M.L. Checa, P. García-Atance, S. Dunner, J. Cañón

Laboratorio de Genética, Dpto. de Producción Animal, Universidad Complutense de Madrid, España

RESUMEN

Con el objetivo de estudiar la diversidad genética existente en un conjunto de razas equinas españolas y el aporte de cada raza a la diversidad global de cara a la aplicación de políticas de gestión de recursos genéticos, se comparan en este artículo dos procedimientos alternativos. El propuesto por Weitzmann se basa en distancias genéticas y su medida de diversidad posee propiedades teóricas que hacen de ella una medida robusta frente a nuevas aportaciones poblacionales y variaciones en las distancias. Produce además un árbol jerárquico de fácil interpretación en lo que a aportaciones genéticas de cada raza se refiere. El análisis de correspondencias es un método estadístico clásico que aplicado en este contexto permite que un concepto análogo al de la tasa de inercia se equipare al de la diversidad aportada por cada raza, además de generar una representación gráfica de las relaciones entre las razas.

Palabras clave: Diversidad genética, Procedimiento de Weitzman, Análisis de Correspondencia.

SUMMARY

In order to study the genetic diversity in a set of spanish horse breeds and the contribution of each breed to the whole diversity, with a posterior use in genetic resources management policies in mind, two different approaches are compared in this paper. Weitzmann's is based upon genetic distances, and the theoretical properties associated to its measure of diversity make it a robust measure with respect to breed additions and variations in the distances. Furthermore, it produces an easily interpretable hierarchical tree for individual breeds genetic contributions studies. The analysis of correspondences is a clasical statistical method which, applied to this context, allows an analogous concept to that of inercy to be translated into the diversity added by each breed. Appart from that, it generates a plot of the relationship among the breeds.

Key words: Genetic diversity, Weitzman's approach, Analysis of correspondence.

Introducción

El término "diversidad genética" aparece, como es lógico, con mucha frecuencia en el contexto de conservación de recursos genéticos. Generalmente aceptamos que la tasa de evolución en poblaciones naturales está limitada por la cantidad de variabilidad genética (Teorema Fundamental de la Selección Natural de Fisher, 1930). No resulta extraño, por lo tanto, que muchos de los esfuerzos en los programas de gestión de recursos genéticos se hayan dedicado al mantenimiento de la diversidad genética, aunque no parece claro que, en el contexto de las especies de animales domésticos, este factor tenga un impacto similar sobre la probabilidad de extinción de una determinada población.

Dependiendo de las características de las poblaciones en estudio y de los objetivos se han propuesto, para definir el concepto de diversidad, diferentes parámetros que combinan la distribución de frecuencias génicas, genotípicas o gaméticas. Entre los parámetros propuestos, los que más frecuentemente aparecen son: el número de loci polimórficos, el número de alelos por locus, la heterocigosis (LEWONTIN, 1974) el Contenido de Información Polimórfica (BOTSTEIN *et al.*, 1980), y con mucha frecuencia, la manera de identificar la excepcionalidad de una población con fines de asignación prioritaria de recursos para su conservación ha sido la utilización de las medidas de distancia genética entre parejas de poblaciones. La capacidad de resolución de estas medidas de distancia no parecen muy evidentes cuando, entre otros problemas, existe la posibilidad de obtener muy diferentes topologías para un mismo conjunto de distancias genéticas y, viceversa, árboles de similar parsimonia originan diferentes estructuras genéticas.

Un aspecto central en programas de conservación genética es la definición de una medida de diversidad y su representación gráfica. El objeto de esta nota es aplicar a un conjunto de razas de la especie equina, dos procedimientos alternativos: a) un procedimiento desarrollado por WEITZMAN (1992, 1993) precisamente para medir la diversidad con fines de su utilización en la gestión de recursos genéticos, b) un análisis de correspondencia que permite estudiar una variable cualitativa (raza) en función de otras variables cualitativas (alelos) y en el que el término clásico de inercia puede ser asimilado al de diversidad.

Métodos de cálculo

Procedimiento de Weitzman

WEITZMAN define una medida de diversidad a partir de, por ejemplo, distancias genéticas, con una serie de propiedades empíricas. Utilizando una notación similar a la de WEITZMAN (1992), supongamos que S representa el conjunto de poblaciones (especies, razas, etc) y que la medida de la distancia genética entre una pareja (i, j) de esas poblaciones, d(i,j) satisface las condiciones: a) $d(i,j) \geq 0$, b) $d(i,i) = 0$, c) $d(i,j) = d(j,i)$. La medida de diversidad V, del conjunto S de poblaciones, V(S) es la solución de:

$$V(S) = \max_{i \in S} \{ V(S \backslash i) + d(i, S \backslash i) \} \quad [1]$$

donde V(S\i) denota la diversidad del conjunto de poblaciones S tras la eliminación de i.

WEITZMAN (1992) considera razonable imponer a la función de diversidad las siguientes condiciones:

• *Monotonicidad de las poblaciones:* cuando se añade un elemento j al conjunto de poblaciones Q, $Q \subseteq S$ la diversidad Q se

incrementa al menos en la cantidad d(j,Q), donde d(j,Q) representa la distancia del elemento j a Q.

$$V(Q \cup j) \geq V(Q) + d(j,Q)$$

• *La propiedad del enlace*: existe al menos una población j que satisface

$$V(S) = V(S\backslash j) + d(j,S\backslash j)$$

Es decir, existe alguna población cuya eliminación reducirá la diversidad en la misma magnitud que la distancia de esa población a la más próxima. Esta propiedad garantiza una conexión natural entre la diversidad medida y las distancias genéticas en las que esa diversidad se basa.

• *Propiedad de los idénticos*: añadir una población idéntica a otra existente en S no proporciona ningún incremento de la diversidad.

• *Continuidad de las distancias:* pequeños cambios en las distancias genéticas dan lugar a pequeños cambios en la diversidad.

• *Monotonicidad de las distancias*: si se incrementan todos los pares de distancias también se producirá un incremento de la diversidad.

En lugar de utilizar [1] para calcular V(S), lo que supone n! cálculos, Weitzman prueba que resolver [1] es equivalente a resolver:

$$V(S) = d[g(S),h(S)] + \max\{V[S\backslash g(S)], V[S\backslash h(S)]\}$$

En la que el elemento g(S) es lo que en el cuadro 1 llamamos el elemento enlace y satisface que $g(S) = \max [V(S\backslash g), V(S\backslash h)]$ y h(S) se denomina representante en esa misma tabla.

Representación gráfica

La información contenida en las distancias genéticas puede representarse gráficamente en forma de un árbol jerárquico cuya topología es única si identificamos las poblaciones que actúan de enlace. Este árbol así construido permite una fácil visualización del efecto a que daría lugar en términos de diversidad la pérdida de una población o un conjunto de ellas.

Debemos tener en cuenta que este árbol es una representación de la diversidad observada a través de las distancias genéticas y, por lo tanto, matrices de distancias genéticas escasamente correlacionadas darán lugar a topologías muy diferentes. Además, también debemos ser conscientes de que, salvo raras excepciones, la topología observada no puede ser interpretada como el reflejo del verdadero proceso evolutivo sufrido por las poblaciones objeto de estudio.

Análisis de correspondencia

El análisis se efectúa a partir de una tabla de contingencia por raza en el que las filas están constituidas por la variable a explicar (la raza) y las columnas el conjunto de variables explicativas (los alelos). El análisis se trata de la modalidad de correspondencia múltiple en el que los diferentes loci se consideran no ligados. Los resultados nos permiten obtener la importancia relativa (inercia) tanto de las razas como de los alelos, una representación gráfica mediante un sistema de puntos situados en un espacio euclidiano y la verificación de su grado de robustez.

Material utilizado

Poblaciones de caballos celtas

Las poblaciones, marcadores genéticos y las medidas de distancia genética que han

sido utilizados se encuentran descritos en CAÑON *et al.* (2000).

Resultados y discusión

La matriz de distancias genéticas puede ser transformada en otra de distancias con propiedades ultramétricas con la ventaja de poder ser utilizadas por programas estándares (ej. NTSYS, RHOLF, 1988) para la construcción de la única topología posible. Uno de los inmediatos aspectos de interés de esta representación gráfica es la visualización de la reducción que se produce en diversidad cuando alguna población o un conjunto de ellas desaparece.

Si observamos la figura 1, en la que se representa la situación de los caballos españoles tipo pony podemos comprobar que, desde la perspectiva de pérdida de diversidad, la situación de las dos poblaciones de las islas Baleares es la más crítica ya que, en el supuesto de que se perdiera la raza mallorquina, la diversidad se vería reducida en un 22%, mayor porcentaje que la desaparición de cualquier otra de las razas incluidas en el análisis, incluso superior al supuesto en el que las dos razas más próximas entre si, el pottoka y el caballo gallego, desaparecieran. Pero la gravedad sería mucho mayor si desaparecieran las dos razas de las islas Baleares ya que ese cluster que desaparece representa el 50% de la diversidad global, es decir, de la diversidad del conjunto de las razas incluidas en este estudio. La situación es completamente diferente si nos fijamos, por ejemplo, en las poblaciones de pottoka y caballo gallego ya que su posición indica que la desaparición de cualquiera de ellas representa una escasa pérdida de diversidad, 9 y 7% respectivamente; incluso la desaparición de ambas razas no resulta tan crítico como la desaparición de una de las de las islas baleares, un 18% frente al 50%. La pérdida de variabilidad genética cuando se elimina un subconjunto de razas no necesariamente tiene que ser igual a la suma de las pérdidas

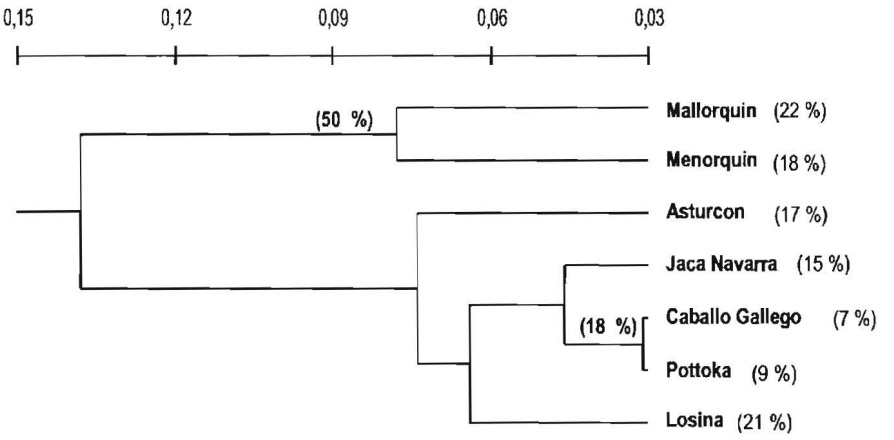


Figura 1. Dendrograma representando la posición de diferentes razas de caballos de origen celta. Las cifras entre paréntesis al lado de cada raza representa el porcentaje de diversidad que se pierde por la desaparición de esa población. Las cifras en los nodos representan el porcentaje de pérdida de diversidad por la eliminación de ese cluster

de cada una de las razas individualmente consideradas que constituyen ese cluster. La aditividad se cumple cuando las razas son independientes.

Otra aplicación de interés resulta de la posibilidad de conocer el rendimiento, en términos de diversidad marginal, que proporciona la inversión en mejorar la probabilidad de supervivencia de una población. En nuestro caso lo que hemos hecho es calcular el incremento marginal en diversidad como consecuencia del incremento de la heterocigosis. Hablar de probabilidad de extinción en poblaciones no naturales como es el caso de las razas de animales domésticos no resulta fácil, pero podríamos pensar en que las acciones de conservación de recursos genéticos animales tengan como uno de sus objetivos tratar de incrementar el nivel de heterocigosis o, más propiamente, reducir el incremento en el nivel de homocigosis, suponiendo que una situación de homocigosis es sinónimo de extinción desde el punto de vista genético. Podríamos, por lo tanto, analizar el rendimiento de la inversión en las poblaciones en función del nivel de homocigosis en la que se encuentran. Para ello hay que utilizar una función que transforme homocigosis en probabilidad de extinción. En ausencia de otra mejor se ha elegido la transformación directa de homocigosis a probabilidad de extinción, es decir, P[extinción] (raza k) = Homocigosis (raza k).

Otro de los aspectos de interés señalado por Weitzman es la posibilidad de calcular la elasticidad de la diversidad genética en las diferentes poblaciones, es decir, en qué medida cambios porcentuales en los niveles de homocigosis provocan cambios más o menos elevados en la diversidad. Una elasticidad elevada indicaría que cambios relativamente pequeños en los niveles de heterocigosis proporcionarían una elevada respuesta en la diversidad del conjunto de las poblaciones.

Recientemente se han propuesto otros procedimientos para medir la diversidad como el Análisis de Correspondencia (LEBART, MORINEAU & WARWICK, 1984) en el que la medida de la 'inercia' de una variable (raza) puede ser equiparable a la diversidad que aporta al conjunto de razas analizadas (LALÖE *et al.*, 1999). Será interesante comprobar en que medida ambos pro-

Cuadro I
Resultados sobre la diversidad acumulada y el orden de aparición de las razas en el dendrograma

Representante	Enlace	Distancia	Variabilidad acumulada	Orden
Pot	Gal	.0310	.0310	Mal
JN	Pot	.0460	.0770	Men
Los	JN	.0640	.1410	Ast
Los	Ast	.0740	.2150	JN
Mal	Men	.0780	.2930	Gal
Mal	Los	.1380	.4310	Pot
				Los

Pot: Pottoka; JN: Jaca Navarra; Los: Losino; Mal: Mallorquina; Men: Menorquina; Gal: Caballo Gallego; Ast: Asturcón.

cedimientos proporcionan resultados parecidos.

El análisis de correspondencia que llevamos a cabo daba como resultado que la raza de caballos asturcones, con un 22%, era la que mayor inercia aportaba, y la población de pottoka la que menos (8%). El primer eje (explica un 31% de la inercia total) claramente diferenciaba al asturcón del resto de las razas. El eje 2 (explica un 24% de la inercia total) a la jaca navarra y el tercer eje (explica un 19% de la inercia total) discriminaba al losino (figura 2). El hecho de que un único marcador (HMS6) contribuyera a las dos primeras componentes con un porcentaje relativamente elevado, el alelo HMS6-171 un 13% al eje 1 y un 11% al eje 2, y los alelos HMS6-161 y HMS6-090 con un 9 y 3% al segundo eje, nos llevó a realizar el análisis después de eliminar este marcador. Los resultados de las contribuciones a la inercia reflejan la misma situación para todas las razas y los ejes discriminan de la misma manera que en el caso anterior.

En el cuadro 2 se presentan las inercias aportadas por cada una de las razas, así como los valores de la diversidad obtenidos por el procedimiento de WEITZMAN. La correlación entre ambos resultados es aproximadamente de 0,6.

La utilización de la medida de distancia genética como argumento base de la medida de diversidad mediante el procedimiento propuesto por WEITZMAN lleva a preguntarnos por la dependencia de los resultados en función del tipo de medida que pueda ser calculada. A pesar de que THAON D'ARNALDI, FOULLEY and OLLIVER (1998) con razas bovinas encuentran valores muy parecidos cuando utilizan dos distancias genéticas (Nei y Cavalli-Sforza) resulta evidente que cambios relativos importantes entre las dis-

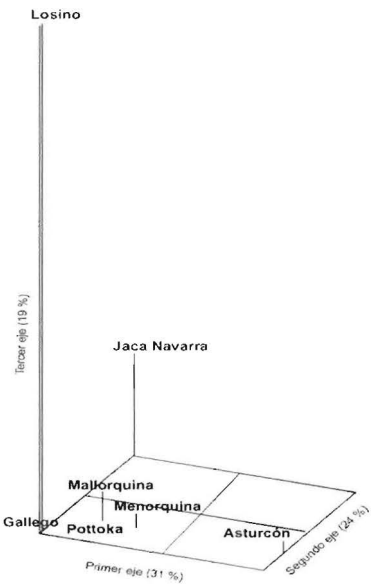


Figura 2. Proyección de las razas sobre los tres primeros ejes del análisis de correspondencia

Cuadro 2
Valores de inercias (análisis de correspondencia) y diversidad (procedimiento de Weitzman) aportadas por cada una de las razas analizadas

	Asturcón	Jaca Nava.	Losino	Mallorquín	Menorquín	Gallego	Pottoka
Inercia	22	17	16	13	15	10	8
Inercia[a]	25	17	17	12	12	10	6
Weitzman	17	15	21	22	17	7	9

([a]) Resultados después de eliminar el marcador HMS6.

tancias genéticas de las parejas de poblaciones pueden proporcionar ordenaciones muy diferentes. En nuestro caso sólo las distancias de FST (WRIGHT, 1965) y de REYNOLDS *et al.* (1983) dieron lugar exactamente al mismo orden, aunque no exactamente a la misma topología, no existiendo correlación distinta de 0 cuando se comparaban los resultados obtenidos con estas distancias y los obtenidos al emplear distancias calculadas bajo el modelo de mutación paso a paso como la distancia cuadrática media (GOLDSTEIN *et al.*, 1995a) o la distancia delta-mu (GOLDSTEIN *et al.*, 1995b) que han sido propuestas para tener en cuenta la elevada tasa de mutación que presentan los microsatélites.

Algunas de las críticas que se hacen a este procedimiento han sido señaladas por THAOM D'ARNOLDI, FOULLEY and OLLIVIER (1998); entre ellas la de no tener en cuenta la variabilidad dentro de poblaciones, algo que sí ocurre con el análisis de correspondencia. Habrá que llevar a cabo estudios adicionales para ver cual es el grado de robustez de los resultados que proporcionan estos métodos cuando se modifican determinadas variables como el tamaño de población, el número de marcadores y su potencia de discriminación, medida de distancia genética utilizada, etc. Otro problema que obviamente no puede ser objetivo de estos métodos es el que concierne a las necesidades de conservación basadas en argumentos o valores no tangibles, como los culturales.

Agradecimientos

El material utilizado en esta presentación es consecuencia de un proyecto financiado por la CICYT (AGF95-0643) y CajaAstur.

Bibliografía

BOSTSTEIN D., WHITE R.L., SKOLNIK M., DAVIS R.W., 1980. Construction of a genetic linkage map in man using restriction frgament length polymorphism. American Journal of Human Genetics, 32, 314-331.

CAÑON J, CHECA ML., CARLEOS C., VEGA-PLA JL., VALLEJO M., DUNNER S., 2000. The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data. Animal Genetics, 31, 39-48.

FISHER R.A., 1930. The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

GOLDSTEIN D.B., RUIZ LINARES A., CAVALLI-SFORZA L.L., FELDMAN M.W., 1995a. An evaluation of genetic distances for use with microsatellite loci. Genetics, 139: 463-471.

GOLDSTEIN D.B., RUIZ LINARES A., CAVALLI-SFORZA L.L., FELDMAN M.W., 1995b. Genetic absolute dating based on microsatelites and the origin of modern humans. Proceedings of National Academy of Sciences USA., 92: 6723-6727.

LALOË D., MOAZAMI-GOUDARZI K., SOUVENIR ZAFINDRAJAONA P., 1999. Analyse des correspondances et biodiversité dans les races domestiques. Société Française de Biométrie, 20 mai 1999, Grenoble, 5 pages.

LEBART L, MORINEAU A, WARWICK K, 1984. Multivariate Descriptive Statistical Analysis. J. Wiley, New York.

LEWONTIN R.C., 1974. The Genetic Basis of Evolutionary Change. Columbia U.P.

REYNOLDS J., WEIR B.S., COCKERHAM C.C., 1983. Estimation of the coancestry coefficient: Basis for a short-term genetic distance. Genetics 105, 767-79.

ROHLF F.J., 1988. NTSYS-pc: Numercial Taxonomy and Multivariate Analysis System. Exeter Software. Heritage Lane. Setauket, New York, USA.

THAON D'ARNOLDI C., FOULLEY J.L., OLLIVIER L., 1998. An overview of the Weitzman approach to diversity. Genetics, Selection and Evolution, 30, 149-161.

WEITZMAN M., 1992. On diversity. Quarterly Journal of Economics, 107, 363-405.

WEITZMAN M., 1993. What to preserve? An application of diversity theory to crane conservation. Quarterly Journal of Economics, 108, 157-183.

WRIGHT S., 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19, 395-420.

ITEA (2000), Vol. 96A N.° 3, 290-295

SOBRE LA CONVERGENCIA DEL MUESTREO DE GIBBS EN LOS MODELOS BAYESIANOS JERÁRQUICOS

L.A. García Cortés, C. Cabrillo

RESUMEN

Desde la introducción del muestreo de Gibbs en mejora genética se ha extendido su uso en combinación con los modelos bayesianos jerárquicos. Aquí mostramos cómo el modelo aumentado con los valores aditivos de los animales y resuelto con un muestreo de Gibbs ofrece resultados que numéricamente no concuerdan con los obtenidos a partir de un modelo bayesiano en el que no se haya realizado el aumento. La aparente paradoja se resuelve en favor de la técnica de aumento de datos y en contra del muestreo de Gibbs cuya convergencia presenta problemas cerca del origen de varianzas.

Introducción

Durante los últimos años, el muestreo de Gibbs (GELFAND y SMITH, 1989) se ha popularizado en mejora genética debido a su sencillez de cálculo y su capacidad para proporcionar las distribuciones marginales de los parámetros de interés. El modelo bayesiano jerárquico se adapta perfectamente a este algoritmo, debido a que las distribuciones condicionales necesarias presentan formas conocidas y sencillas de implementar (WANG *et al.*, 1993).

En el modelo bayesiano jerárquico, es bien conocido que las distribuciones condicionales de la varianza entre machos presentan una forma proporcional a una ji-cuadrado inversa. No parece extraño entonces que las distribuciones marginales de la varianza entre machos obtenidas usando el muestreo de Gibbs, es decir, promediando condicionales a partir de una cadena de Markov, presenten una ordenada nula en el cero. Tal es el caso de las figuras presentadas en (KORSGAARD *et al.*, 1998; Sorensen *et al*, 1994; STRANDEN y GIANOLA, 1999; WANGET *et al.*, 1993) entre otras.

Por contra, en los resultados que presentamos en GARCÍA CORTÉS *et al.* (1999), la distribución marginal de un componente de varianza toma un valor mayor que cero en el origen. La diferencia entre este trabajo y los anteriormente mencionados es únicamente la manera de implementar el muestreo de Gibbs. Concretamente en este último trabajo se utiliza un modelo donde se integran los efectos aditivos de los animales, es decir, un análisis en el que no se utiliza un modelo aumentado.

En este último trabajo se construye la verosimilitud a partir de una única distribución, esto es, asumiendo que los datos observados siguen una distribución normal multivariante cuyos parámetros se desean inferir. En el modelo que vamos a usar aquí (un modelo macho por simplicidad) esta distribución se expresa como

$$y \sim N\left(Xb, ZZ'\sigma_u^2 + I\sigma_e^2\right) \qquad [1]$$

Como se expone en lo que sigue, las distribuciones de la componente de varianza

presentan diferencias sustanciales en un entorno del cero según que en el análisis se utilize o no la técnica de aumento de datos en clara contradicción con la esencia misma de dicha técnica, a saber, que la inferencia no se ve modificada en el proceso (TANNER y WONG, 1987). Demostraremos aquí como la aparente paradoja se debe a un problema de convergencia del muestreo de Gibbs cuando se aplica al modelo con aumento de datos. Como cabría esperar, la inferencia en sí es independiente del modelo elegido y, en contra de la intuición, las verdaderas distribuciones marginales no son nulas en el origen aunque de una forma singular que impide la convergencia correcta del muestreo de Gibbs.

El modelo

En aras de una mayor simplicidad y, puesto que no afecta a la generalidad de nuestras conclusiones, nos restringiremos a un modelo macho descrito por

$$y = Xb + Zu + e$$

donde $y \sim N\left(Xb, ZZ'\sigma_u^2 + I\sigma_e^2\right)$ son los datos observados, b son efectos fijos desde el punto de vista frecuentista,

$$u \sim N\left(0, I\sigma_u^2\right)$$

son los efectos genéticos de los padres,

$$e \sim N\left(0, I\sigma_e^2\right)$$

son los residuos del modelo y X y Z son las matrices de incidencias que relacionan los niveles de cada efecto con las observaciones.

La distribución posterior de σ_u^2 puede obtenerse por marginalización con respecto a b y σ_e^2 de la posterior conjunta:

$$f\left(b, \sigma_u^2, \sigma_e^2 \mid y\right) \qquad [2]$$

Llamaremos a éste el análisis 1 (o modelo 1). Se corresponde con el utilizado en GARCIA CORTÉS *et al.* (1999).

Es bien conocido que puede simplificarse el problema aumentando el modelo anterior con las variables incluidas en u, para después obtener la marginal posterior de σ_u^2 marginalizando con respecto a u, b y σ_e^2 en:

$$f\left(u, b, \sigma_u^2, \sigma_e^2 \mid y\right) \qquad [3]$$

Este segundo análisis (o modelo) está basado en el aumento de datos y está garantizado que las marginales de σ_u^2 en [2] y [3] son iguales (TANNER y WONG, 1987). Este último será el análisis 2 y consiste en el modelo bayesiano jerárquico habitual en mejora genética (WANG *et al.*, 1993).

Implementación

Las marginalizaciones en [2] y [3] van a ser realizadas usando un muestreo de Gibbs. Dicha marginalización se llevará a cabo de la forma habitual en tres fases: definir la posterior conjunta en función de la verosimilitud y distribuciones *a priori*, tomar la condicional de σ_u^2 en cada ciclo de la cadena y promediar las condicionales. Es decir, no usamos los puntos muestreados sino que acumulamos las densidades condicionales, método conocido como de Rao-Blackwell (GELFAND y SMITH, 1989). En ambos análisis las posteriores son (suponemos todas las priori planas):

Análisis 1

$$f\left(b, \sigma_u^2, \sigma_e^2 \mid y\right) \propto f\left(y \mid b, \sigma_u^2, \sigma_e^2\right)$$

$$f\left(b, \sigma_u^2, \sigma_e^2 \mid y\right) \propto \left|ZZ'\sigma_u^2 + I\sigma_e^2\right|^{-0.5}$$

$$\exp\left\{-0,5(y - Xb)'\left(ZZ'\,\sigma_u^2 + I\sigma_e^2\right)^{-1}(y - Xb)\right\}$$

Análisis 2

$$f\left(u,b,\sigma_u^2,\sigma_e^2 \mid y\right) \propto f\left(y \mid u,b,\sigma_u^2,\sigma_e^2\right) f\left(u \mid \sigma_u^2\right)$$

$$f\left(u,b,\sigma_u^2,\sigma_e^2 \mid y\right) \propto \sigma_e^{-n}\sigma_u^{-q}$$

$$\exp\left\{-\frac{(y - Xb - Zu)'(y - Xb - Zu)}{2\sigma_e^2} - \frac{u'\,u}{2\sigma_u^2}\right\}$$

donde n es el número de datos y q es el número de animales.

Tomando las condicionales completas para σ_u^2 en ambos modelos se obtiene

Análisis 1

$$f\left(\sigma_u^2 \mid b,\sigma_e^2,y\right) \propto \left|ZZ'\,\sigma_u^2 + I\sigma_e^2\right|^{-0.5}$$

$$\exp\left\{-0,5(y - Xb)'\left(ZZ'\,\sigma_u^2 + \sigma_e^2\right)^{-1}(y - Xb)\right\}$$

Análisis 2

$$f\left(\sigma_u^2 \mid u,b,\sigma_e^2,y\right) \propto \sigma_u^{-q}\exp\left\{-\frac{u'\,u}{2\sigma_u^2}\right\} \qquad [4]$$

en donde las correspondientes constantes de proporcionalidad son independientes de u. Una vez obtenidas las condicionales completas, las marginales se obtienen promediando las de cada ciclo del muestreo de Gibbs. Centrándonos en el punto $\sigma_u^2 = 0$, las densidades a promediar resultan en

Análisis 1

$$f\left(\sigma_u^2 = 0 \mid b,\sigma_e^2,y\right) \propto \left|I\sigma_e^2\right|^{-0.5}$$

$$\exp\left\{-0,5(y - Xb)'\left(I\sigma_e^2\right)^{-1}(y - Xb)\right\} \qquad [5]$$

Análisis 2

$$\lim_{\sigma_u^2 \to 0} f\left(\sigma_u^2 \mid u,b,\sigma_e^2,y\right) \propto \delta(u'\,u) \qquad [6]$$

donde $\delta(\ )$ denota delta de Dirac. La variable u'u representa una densidad localizada en cero. Es fácil convencerse de que este comportamiento singular es el correcto inspeccionando [4]. En el caso de que u'u valga cero el factor exponencial en el lado derecho de [4] se reduce a 1 de tal forma que la divergencia en $\sigma_u^2 = 0$ del prefactor σ_u^{-q} no se ve compensada en este caso por dicho factor exponencial. Al mismo tiempo, puesto que [4] representa una distribución de probabilidad es fácil convencerse de que el límite $\sigma_u^2 \to 0$ aunque divergente, se aproxima manteniendo el area finita con respecto a todas las variables y en particular con respecto a u'u. Matemáticamente este tipo de divergencia define la delta de Dirac que, si bien no es una función en el sentido habitual, adquiere un sentido riguroso bien definido dentro de la llamada teoría de las funciones generalizadas. En particular, la integral de la delta para este caso en el que está centrada en el origen del intervalo de integración ha de identificarse con 1/2. Este tipo de comportamiento singular conduce a una probabilidad finita en $\sigma_u^2 \to 0$ garantizando la equivalencia con el modelo 1 cuya probabilidad finita en $\sigma_u^2 = 0$ es evidente por simple inspección de [5]. Sin embargo, al mismo tiempo este comportamiento singular impide una convergencia correcta del muestreo de Gibbs puesto que la probabilidad de alcanzar el valor u'u = 0, el único punto que contribuye al valor no nulo de la marginal en el origen, es infinitesimal.

Ejemplo numérico

Dejaremos un análisis matemático más completo para una publicación posterior.

Aquí nos limitaremos a ilustrar el problema con un ejemplo numérico. La muestra corresponde a un modelo con 50 machos, 25 efectos fijos en un solo factor y 500 datos. Para la simulación se usó una varianza residual de 95 y una varianza entre machos de 5. La figura 1 presenta las marginales en ambos modelos y corrobora lo anteriormente expuesto: Ambas condicionales son distintas en el punto $\sigma^2_u = 0$, forzando a que las marginales sean distintas en el entorno de cero.

Un simple experimento numérico puede convencernos de que el resultado correcto es el del análisis 1. Para ello utilizaremos una *a priori* discreta para σ^2_u, con el 50% de la densidad en el punto 0.1 y el 50% restante en el punto 0,2 (dos puntos escogidos por estar en el entorno de cero) y comparemos los resultados de los siguientes casos:

Caso 1. Densidad marginal posterior para el análisis 1 con una *a priori* plana.

Caso 2. Densidad marginal posterior para el análisis 2 con una *a priori* plana.

Estos dos primeros casos se obtienen directamente de la figura 1, tomando las ordenadas en los puntos 0,1 y 0,2.

Caso 3. Probabilidad marginal posterior para el análisis 1 con una *a priori* discreta.

Caso 4. Probabilidad marginal posterior para el análisis 2 con una *a priori* discreta.

Estos dos últimos casos se obtienen repitiendo el análisis con dicha *a priori* incluida explícitamente en el modelo. La diferencia con los dos análisis anteriores está en el muestreo de σ^2_u: esta condicional solamente tiene probabilidad en los puntos 0,1 y 0,2 y la cadena muestrea solamente entre estas dos

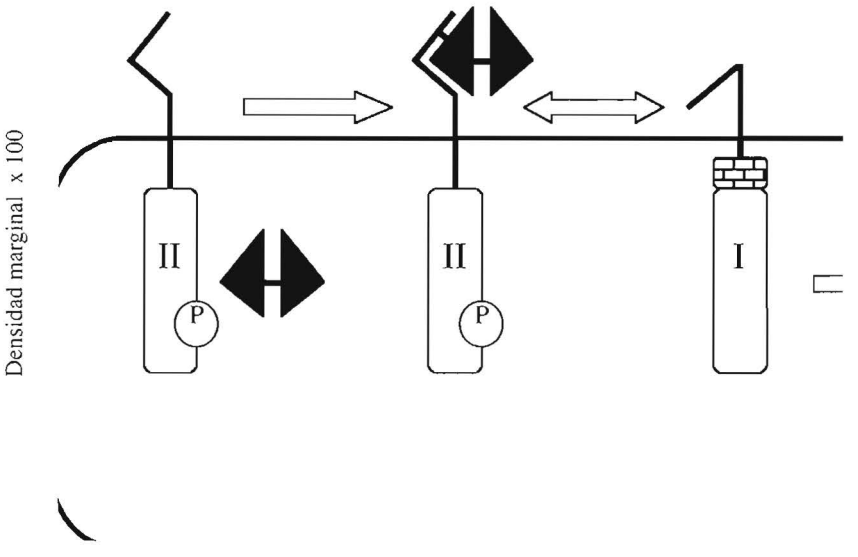


Figura 1. Distribuciones marginales posteriores para la varianza entre machos en los modelos 1 y 2

Cuadro 1
Densidades o probabilidades marginales posteriores en loscuatro casos

Caso	Análisis	*A priori*	Densidad en 0,1	Densidad en 0,2	Cociente
1	1	Plana	0,06572	0,06948	1,057
2	2	Plana	0,01905	0,05441	2,856
3	1	Discreta	0,4861	0,5139	1,057
4	2	Discreta	0,4822	0,5178	1,073

posibilidades. Nótese que en estos dos casos hablamos de probabilidades en lugar de densidades, ya que la distribución *a priori* discreta da lugar a una marginal posterior discreta. La idea subyacente es que a pesar de que las *a priori* discretas usadas están dentro de la zona en la que ambos modelos discrepan el carácter localizado de éstas debería evitar, o al menos disminuir fuertemente, la influencia perniciosa de la mala convergencia en el origen. Por otro lado, como ambas *a priori*, la plana y la discreta, presentan el mismo cociente entre sus valores en los dos puntos estudiados (esto es, 1 en ambos casos), los correspondientes cocientes de las posteriores deberían también coincidir (aunque, evidentemente, ya no serían igual a 1).

El cuadro 1 confirma estas hipótesis dentro de una precisión numérica más que razonable. Efectivamente, los casos 1 y 3 correspondientes al modelo 1 dan el mismo cociente de densidades. Así mismo, los casos 3 y 4 correspondientes a modelos diferentes pero con la *a priori* discreta coinciden razonablemente bien tanto en las densidades como en el cociente. El cociente en el caso 2, por contra, difiere fuertemente del obtenido en los otros casos, confirmando así, que los dos modelos son equivalentes pero que el mues-

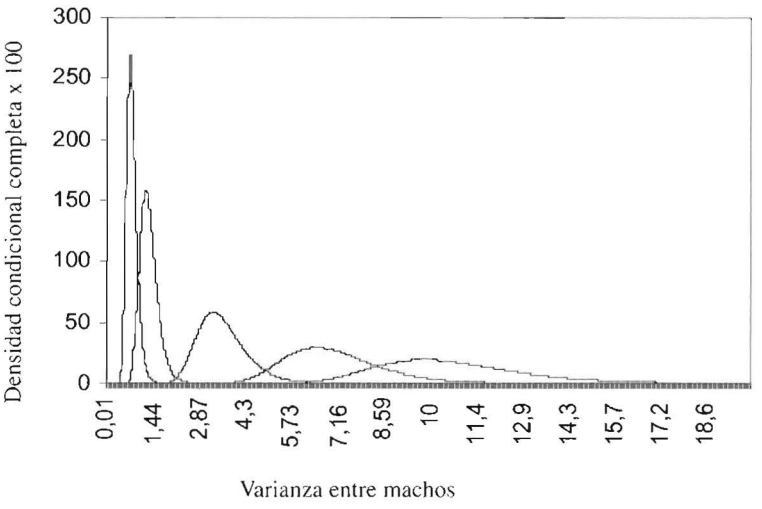


Figura 2. 5 condicionales de las que se promedian para obtener la gráfica anterior

treo de Gibbs es inapropiado cuando las *a priori* toman valores finitos en el origen.

La figura 2 permite visualizar el comportamiento patológico que invalida el muestreo de Gibbs. En ella se muestra una serie de 5 condicionales de las que se promedian para obtener la curva del análisis 2 de la figura 1 correspondientes de derecha a izquierda a valores decrecientes de u'u. Según u'u se va aproximando a cero, se observa claramente como al mismo tiempo que se van desplazando hacia el origen se van estrechando manteniendo un área finita. En el límite u'u = 0 colapsan en una condicional infinitamente estrecha, centrada en el origen y de área finita, esto es, en algo proporcional a una delta de Dirac. A la vista de la figura 2 es fácil entender que según nos acercamos al origen la densidad marginal se hace progresivamente más difícil de estimar puesto que en el promedio van pesando cada vez más, condicionales más estrechas y más infrecuentes, hasta que en el propio origen la estimación fracasa completamente.

Por último, nótese que el haber utilizado distribuciones *a priori* no informativas (planas) y por tanto impropias, no tiene relación alguna con el mal comportamiento del muestreo. De hecho, en las simulaciones numéricas las *a priori* planas están perfectamente definidas puesto que el rango de la variable aleatoria es evidentemente finito. Se han utilizado nada más que por su simplicidad de implementación. Cualquier otra *a priori* con valor no nulo y finito en el origen presentaría el mismo problema de convergencia.

Implicaciones

La implicación de lo aquí expuesto es clara: En caso de necesitar probabilidades de intervalos en la cola izquierda de la distribución marginal posterior, se debería o bien prescindir del muestreo de Gibbs junto con aumento de datos o bien corroborarlo con alguna simulación similar a la expuesta en la sección anterior. En el caso extremo en el que se necesite la ordenada en el eje, el modelo jerárquico es del todo inadecuado.

Así mismo, habría que ser cuidadoso con los análisis que proporcionan marginales cuya densidad disminuye rápidamente al acercarse al cero. Estos deberían ser revisados, bien utilizando un experimento similar al empleado en la sección anterior (usando una *a priori* discreta) o bien implementando un modelo sin aumento de datos.

Bibliografía

GARCÍA CORTÉS *et al.*, 1999. ITEA, Vol Extra.

GELFAND Y SMITH, 1990. J. AMER. Stat. Assoc., 85:398-409.

KORSGAARD *et al.*, 1998. Genet. Sel. Evol. 30:241-256.

SORENSEN *et al.*, 1994. Genet. Sel. Evol. 26:333-360.

STRANDEN Y GIANOLA, 1999. Genet. Sel. Evol. 31:25-42.

TANNER Y WONG, 1987. Tools for statistical inference. Springer.

WANG *et al.*, 1993. Genet. Sel. Evol. 25:41-62.

ITEA (2000), Vol. **96A N.° 3,** 296-301

PROPUESTA DE ESTIMACIÓN DE PARÁMETROS GENÉTICOS DEL CARÁCTER ÍNDICE DE CONVERSIÓN EN DOS LÍNEAS DE CONEJOS SELECCIONADAS POR CRECIMIENTO

E.A. Gómez*,, O. Rafel*, J. Ramón*, O. Perucho***

* Unitat de Cunicultura. IRTA, Granja Escuela Torre Marimon, 08140 Caldes de Montbui, Barcelona, España

** Departamento de Producción Animal y Ciencia de los Alimentos, Univ. Cardenal Herrera CEU. Ed. Seminario. 46113 Moncada, Valencia, España

Introducción

El conocimiento preciso de los parámetros genéticos es más necesario en procesos de selección multicarácter, pues el progreso genético es sensible a la precisión de la matriz de varianzas covarianzas genética (MEYER y HILL, 1983). Pese a la importancia económica del carácter índice de conversión en cunicultura (ARMERO y BLASCO, 1992), existe poca información sobre su relación genética con el carácter utilizado en los Centros de Selección, la velocidad de crecimiento en el período de engorde. No se realiza selección directa sobre el índice de conversión debido al coste que representa el control individual de los gazapos durante el engorde, en tiempo y en instalaciones. La selección indirecta esperada es función directa de la correlación genética entre ambos caracteres, parámetro en el que se centra la presente propuesta. Sin embargo, la estimación de parámetros genéticos, especialmente de las correlaciones, es complicada. Se ha de recurrir a un tipo de diseño que permita realizarla sin aumentar en demasía el tamaño de las bases de datos a analizar.

La selección por velocidad de crecimiento ha tenido éxito. Este carácter se presenta en todos los individuos y la heredabilidad es media alta, lo que permite emplear métodos simples de selección. ESTANY *et al.* (1992) estiman la respuesta genética en dos líneas seleccionadas, con valores de 0,65 y 0,55 g/día por generación. Estas líneas se fundieron en el año 1988, originando la línea Rosa. En la actualidad, esta población (línea *Rosa*) seleccionada en la Universidad Politécnica de Valencia se encuentra en su 21.ª generación, y el carácter de selección sigue siendo la velocidad de crecimiento entre 28 y 63 días. La línea *Caldes* se fundó en el año 1983, y fue seleccionada inicialmente por un objetivo global (RAFEL *et al.*, 1990) que incluía la velocidad de crecimiento además del peso de la camada al destete. Desde 1993 se selecciona únicamente por la tasa de crecimiento entre 32 y 60 días de vida. Se ha estimado una respuesta de 1,06 g/día por año (GÓMEZ *et al.*, 2000).

Las líneas seleccionadas por crecimiento en España presentan menores índices de conversión y mayores velocidades de crecimiento que las líneas seleccionadas por caracteres maternales (TORRES *et al.*, 1992; FEKI *et al.*, 1996; RAMON *et al.*, 1996). TORRES *et al.* (1992) y FEKI *et al.* (1996) indican que la línea *Rosa* presenta un índice

de conversión menor (entre 0,2 y 0,4 puntos) que las líneas *Amarilla* y *Verde*, y una diferencia en velocidad de crecimiento de 8 a 9 g/día mientras que estas líneas no diferían en velocidad de crecimiento al inicio del proceso de selección (BLASCO *et al.*, 1983; ESTANY *et al.*, 1992). Los consumos de pienso fueron un 14% mayores en la Línea Rosa que en las líneas maternales (FEKI *et al.*, 1996). En el trabajo de RAMON *et al.* (1996) la línea *Rosa* fue la de mayor crecimiento (52 g/día), seguida de la línea *Caldes* (45,6 g/día) que presentó una fuerte depresión del crecimiento en el verano. El valor del índice de conversión fue de 2,7 en la Línea *Rosa* y de 2,9 en la línea *Caldes*. GÓMEZ *et al.* (1998b) presentan un resumen de parámetros productivos en el período 96-97, y en el cuadro 1 se presentan las velocidades de crecimiento.

La selección por velocidad de crecimiento parece modificar la curva de crecimiento y conduce al aumento de peso adulto (BLASCO *et al.*, 1990; PILES, 2000) aumentando los pesos a lo largo de la curva (BLASCO *et al.*, 1996), pudiendo llevar a largo plazo a líneas gigantes, con problemas asociados de manejo de los mismos, de adaptación a los suelos de rejilla y de mayores necesidades de mantenimiento.

GÓMEZ *et al.* (1998a) comparan las características de la canal de líneas seleccionadas por caracteres diferentes. En un primer experimento, al comparar al mismo peso de sacrificio (líneas *Amarilla, Rosa* y *Verde*), los animales de la línea Rosa eran 6-7 días más jóvenes, y el rendimiento a la canal era 3 puntos menor. En un segundo experimento (líneas *Caldes, Prat, Rosa* y *Verde*) se realizaron comparaciones a la misma edad. La línea *Rosa* era 500 g más pesada, y la línea *Caldes* 200 g. Las diferencias entre estas dos líneas se redujeron hasta los 130 g entre las canales comerciales, al tener la línea *Caldes* un mayor rendimiento a la canal (58 frente a 56%), similar a las líneas maternales (*Verde* y *Prat*).

La selección por velocidad de crecimiento aumenta el consumo diario. Debido a este aumento del apetito, podía temerse un aumento de los depósitos grasos, como en aves (CHAMBERS *et al.*, 1990) o en cerdos (KHULERS y JUNGST, 1993); a pesar de ser la grasa un tejido tardío. Si el sacrificio se produce a igual peso, y corrigiendo por canal de referencia constante, la línea *Rosa* presentaba el menor depósito graso. En el experimento 2 de GÓMEZ *et al.* (1998a), a igual edad y corrigiendo igualmente por peso de la canal de referencia la línea Rosa también tenía el menor depósito perirrenal, y la línea *Caldes* el mayor. En ninguno de los dos experimentos se observaron diferencias en grasa interescapular.

No existe evidencia directa de la respuesta genética indirecta en índice de conversión. Las estimas publicadas de la correlación genética entre velocidad de crecimiento e

Cuadro 1
Velocidad de crecimiento (según período de engorde) de las líneas de conejo seleccionadas en España

Línea	Amarilla	Hiper	Rosa	Verde	Caldes	Prat
Período		28-63			32-60	
VC (g/día)	36,9	38,1	47,8	37,9	50,2	38,7

índice de conversión se presentan en el cuadro 2 (tomada de BASELGA y BLASCO, 1989). Los resultados parecen poco coherentes. La estimación de la correlación genética entre dos caracteres es compleja, más aún que la estimación de las heredabilidades.

Objetivo

El objetivo es la estimación de los parámetros genéticos de los caracteres velocidad de crecimiento e índice de conversión en dos poblaciones seleccionadas por velocidad de crecimiento (ver figura 1). Se requiere, por tanto, realizar las siguientes actividades:

- Crecimiento e índice de conversión en generación parental.
- Selección de futuros reproductores (selección elíptica).
- Crecimiento e índice de conversión en generación filial.

Diseño

Todos los animales en control se alojarán individualmente. Durante el engorde (32 a 60 días de vida) se realizará un seguimiento diario (bajas, falta de pienso o agua, instalaciones) y un control semanal del peso vivo y del consumo de pienso.

Para optimizar las unidades de alojamiento experimentales se ha recurrido a un diseño que permita mayores precisiones en la estimación de los parámetros genéticos (selección elíptica de CAMERON y THOMPSON, 1986). Este método es eficaz en el caso de poder disponer de los datos de dos generaciones (parental y filial).

Así como en situaciones unicarácter la selección de los animales extremos aumenta la eficacia del diseño, la selección elíptica propone la utilización de un índice cuadrático, eligiendo como reproductores los animales que superen el umbral establecido, que no es más que seleccionar los individuos periféricos de una binormal. La eficacia puede aún aumentarse realizando apareamientos asociativos o dejando un número de hijos variable según el valor de los padres (HILL, 1990).

Se han seguido las indicaciones aclaratorias de BLASCO *et al.* (1990) que se basan en el artículo inicial de CAMERON y THOMPSON (1986). Como parámetros a priori, asumimos unas heredabilidades de 0,34 para ambos caracteres y una correlación genética de –0,7

Cuadro 2
Correlaciones genéticas entre índice de conversión y velocidad de crecimiento

Autores	Año	Período	Estimas
KEIL	1971	42-84	-0,72 (en hembras)
LAMPO y VAN DER BROECK	1975		-1,082
VRILLON *et al.*	1979	24-70	-0,19
RANDI y SCOSSIROLI	1980	56-77	-0,99
		77-100	-1,20
		56-100	-0,77

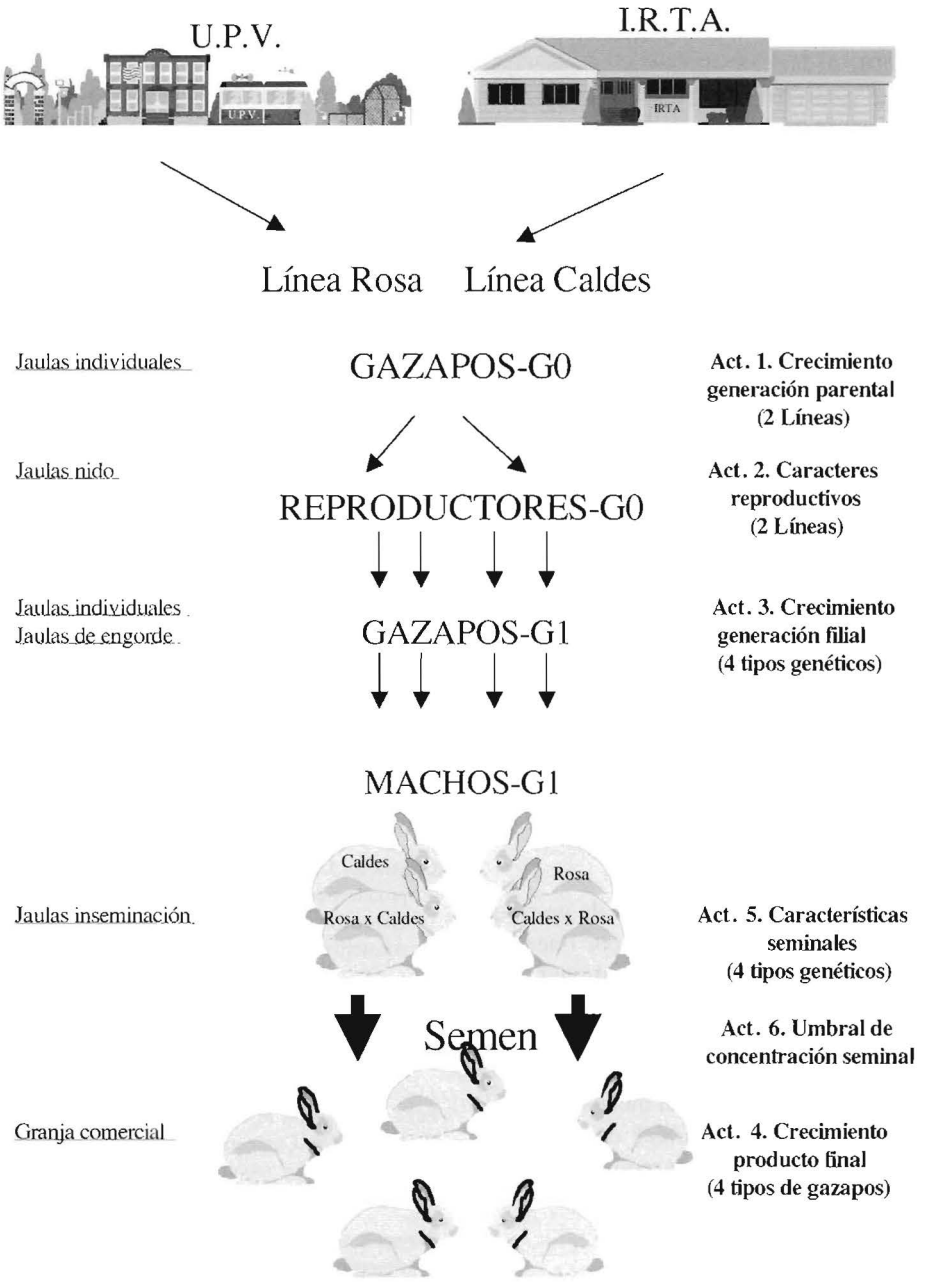


Figura 1. Esquema proyecto de investigación SC00-11: Estimación de parámetros genéticos del carácter índice de conversión en dos líneas especializadas en crecimiento. Estudio de la cinética de producción de semen y el posible interés del cruce dialélico entre líneas seleccionadas

(BASELGA y BLASCO, 1989), y una correlación fenotípica de –0,40 (estimación propia). La matriz de la transformación canónica que permite incorrelacionar fenotípica y genotípicamente los caracteres canónicos es

$$T = \begin{pmatrix} 0,913 & 0,913 \\ 0,598 & -0,598 \end{pmatrix}$$

con lo que las heredabilidades canónicas son 0,17 y 0,41 para la velocidad de crecimiento y el índice de conversión, respectivamente. Así mismo, la heredabilidad canónica intermedia es de 0,267, y el valor intermedio d es igual a 3. Se realizará una presión de selección del 25% en machos y del 33% en hembras. La matriz de varianzas covarianzas de los parámetros genéticos de los caracteres tipificados, con una presión de selección no óptima (0,29) será

$$\begin{pmatrix} 0,021 & -0,008 & 0,003 \\ & 0,012 & -0,008 \\ & & 0,021 \end{pmatrix}$$

El umbral de selección, aplicando la presión óptima sería de –2 log p_E= 1,77. Aplicando una presión no óptima, el umbral será de 1,07.

Se partirá de una población de gazapos con 60 hembras y 40 machos por línea. Se realizará una presión de selección del 25% en machos y del 33% en hembras, dejando 20 madres y 10 padres. Además, serán necesarios al menos 4-5 descendientes medidos por hembra. Se requieren al menos 200 huecos para control individual de consumo (12,5 módulos de jaulas de reposición).

Bibliografía

ARMERO E., BLASCO A., 1992. Economic weights for rabbit selection indices. Jounal of Applied Rabbit Research, 15: 637-642.

BASELGA M., BLASCO A., 1989. Mejora genética del conejo de producción de carne. Ed. Mundi-Prensa. Madrid. 110 pp.

BLASCO A., BASELGA M., GARCÍA-XIMÉNEZ F., 1983. Análisis fenotípico de caracteres productivos en el conejo para carne. I. Caracteres de crecimiento. Archivos de Zootecnia, 32(123), 111.

BLASCO A., GOU P., SANTACREU M.A., 1990a. Diseño de experimentos para estimar parámetros genéticos mediante selección elíptica. V Reunión de Mejora Genética Animal. Córdoba, 23-42.

BLASCO A., GOU P., SANTACREU M.A., 1990b. The effect of selection on changes in body composition of rabbits. 4th World Congress on Genetics Applied to Livestock Production. Edimburgo. Vol XVI. 362-365.

BLASCO A., PILES M., RODRIGUEZ E., PLA M., 1996. Effect of selection for growth rate on the live weight growth curves in rabbits. 6th World Rabbit Congress, Toulouse, Francia. Vol 2: 245-248.

CAMACHO J., 1989. Estimación de correlaciones genéticas entre caracteres reproductivos y de crecimiento en conejos. Tesis Doctoral. Universidad Politécnica de Valencia.

CAMERON N., THOMPSON R., 1986. Design of multivariate selection experiments to estimate genetic parameters. Theoretical and Applied Genetics, 72, 466-476.

CHAMBERS J.R., 1990. Genetics of growth and meat prodduction in chickens. En "Poultry Breeding and Genetics" de Crawford R.D., 1990. Ed. Elsevier, Amsterdam, pp 559-644.

ESTANY J., CAMACHO J., BASELGA M., BLASCO A., 1992. Selection response of growth rate in rabbits for meat production. Genetics, Selection and Evolution, 24, 527-537.

FEKI S., BASELGA M., BLAS E., CERVERA C., GÓMEZ E.A., 1996. Comparison of growth and feed efficiency among rabbit lines selected for different objectives. Livestock Prodcution Science, 45, 87-92.

GÓMEZ E.A., BASELGA M., RAFEL O., RAMON J., 1998a. Comparison of carcass characteristics in five strains of meat rabbit selected on different traits. Livestock Production Science, 55, 53-64.

GÓMEZ E.A., BASELGA M., RAFEL O., GARCÍA M.L., RAMON J., 1998b. Selection, diffusion and performances of six Spanish lines of meat rabbit. 2nd

International Conference of Rabbit Production in Hot Climates. Adana, Turquía.

GÓMEZ E.A., RAFEL O., RAMON J., 2000. Preliminary genetic analyses of Caldes line: a selection experiment for a global objective. 7th World Rabbit Congress, Valencia, España. (in press).

HILL W.G., 1990. En "Advances in statistical methods for genetic improvement of livestock" de Gianola D. y Hammond K. Ed. Verlag.

KEIL J.P., 1971. Variations génétiques de la croissance pondérale et de l'efficacité alimentaire chez le lapin de chair en croissance. Mémoire de fin d'études. E.N.S.A.M., 34 pp.

KHULERS D.L., JUNGST S.B., 1993. Correlated responses in reproductive and carcass traits to selection for 200-day weight in Landrace pigs. Journal of Animal Science, 71, 608-617.

LAMPO P., VAN DER BROECK L., 1975. The influence of heritability of some breeding parameters and the correlation between parameters with rabbits. Journal of Animal Breeding and Genetics, 39(6), 208-211.

MEYER K., HILL H.G., 1983. Journal of Animal Breeding and Genetics, 100, 27-32.

PILES M.M., 2000. Análisis bayesiano de los efectos de la selección sobre las curvas de crecimiento y la calidad dela canal y de la carne de conejo. Tesis Doctoral. Universidad Politécnica de Valencia

RAFEL O., TRAN G., UTRILLAS M., RAMON J., PERUCHO O., DUCROCQ V., BOSCH A., 1990. Sélection pour un objectif global (poids de la portée à 60 jours) en générations chevauchantes dans une lignée synthétique de lapins. Etude de la variabilité non-génétique de la taille et du poids de portée à différents stades. Options Méditerranéennes, A-8, 75-82.

RAMON J., GÓMEZ E.A., PERUCHO O., RAFEL O., BASELGA M., 1996. Fed efficiency and postweaning growth of several Spanish selected lines. 6th World Rabbit Congress, Toulouse, Francia, Vol 2: 351-354.

RANDI E., SCOSSIROLLI R.E., 1980. Genetic analysis of production traits in Italian New Zealand white and California pure.breed populations. II World Rabbit Congres, Barcelona, España, Vol 1. 192-201.

TORRES C., BASELGA M., GÓMEZ E.A., 1992. Effect of weight daily gain selection on gross feed efficiency in rabbits. Journal of Applied Rabbit Research, 15, 872-878.

VRILLON J.L., DONALD R., POUJARDIEU B., ROUVIER R., 1979. Sélection et testage de lapins mâles de croisement terminal 1972-1975. Bulletin Technique Departement Génétique Animale, INRA.

ITEA (2000). Vol. **96A N.° 3,** 302-310

CARACTERIZACIÓN DE VARIANTES GENÉTICAS EN LA κ-CASEÍNA CAPRINA

M. Habib Yahyaoui, A. Coll, A. Sánchez, J.M. Folch

Unitat de Genètica i Millora, Departament de Patologia i Producció Animals, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193 Bellaterra, España

RESUMEN

Se ha analizado el polimorfismo de la k-caseína caprina mediante la técnica *Base Excision Sequence Scanning (BESS)* y secuenciación de ácidos nucleicos. Se han detectado siete posiciones polimórficas, tres de las cuales son silenciosas y las otras cuatro determinan cambio aminoacídico. La asociación entre estos polimorfismos ha sido estudiada y se han identificado tres alelos de la κ-caseína caprina, denominados A, B y C. Se han desarrollado protocolos para el genotipado de la variante C por PCR-RFLP utilizando los enzimas de restricción *Alw*44I y *Bse*NI. La frecuencia de este alelo es muy baja en las razas españolas, pero es más frecuente en la raza francesa Saanen. Serán necesarios estudios adicionales en diferentes poblaciones caprinas para establecer la distribución de estos alelos y determinar sus efectos sobre la calidad y propiedades funcionales de la leche.

SUMMARY

Polymorphisms in the goat k-casein gene were studied using the *Base Excision Sequence Scanning (BESS)* method and sequencing. Seven polymorphic sites were detected, three of these were silent mutations while the other four produced amino acid substitutions. The association between these polymorphisms was investigated, which resulted in the identification of three goat κ-casein alleles, designated A, B, and C. Protocols for rapid genotyping of the C variant were developed by PCR-RFLP using *Alw*44I and *Bse*NI restriction endonucleases. The occurrence of this allele was found to be very low in Spanish breeds but more frequent in the French Saanen goat. Further studies among different goat populations are necessary to establish the distribution of these alleles and their effects on the quality and functional properties of milk.

Introducción

La leche de los rumiantes contiene seis proteínas principales: las caseínas (αs1, αs2, β, y κ) y las proteínas del lactosuero (β-lactoglobulina y α-lactalbúmina). Las cuatro caseínas son el componente mayoritario, constituyendo el 76-86% de la proteína total de la leche (SWAISGOOD, 1992). La κ-caseína (κ-Cn) es esencial para la formación y estabilización de las micelas y tiene un efecto importante sobre las propiedades de la leche. Para la formación del cuajo en la producción del queso, es necesario la digestión de la κ-Cn, produciéndose la precipitación de las micelas. En bovino, se han descrito 6 varian-

tes (A, B, C, E, F, y G) de la κ-Cn (revisado en KAMINSKI *et al.*, 1996). Varios estudios han descrito que el genotipo BB de la κ-Cn determina unas mejores propiedades de la leche para la producción de queso (revisado en NG-KWAI-HANG, 1998). En particular, este genotipo ha sido asociado con un cuajo más firme, una reducción del tiempo necesario para la formación del cuajo y un incremento en el rendimiento en la producción de queso.

La κ-Cn caprina fue aislada de la leche por Zittle y Custer (1966) y su secuencia de 171 aminoácidos fue determinada posteriormente por MERCIER *et al.* (1976a, b). Debido a modificaciones post-traduccionales, la κ-Cn se muestra heterogénea en el análisis por electroforesis, con al menos 5 formas diferentes (ADDEO *et al.*, 1978). La secuencia nucleotídica del ADNc que codifica la κ-Cn caprina y la secuencia promotora del gen han sido aisladas y secuenciadas (COLL *et al.*, 1993, 1995). El gen de la κ-Cn comprende 5 exones, de los cuales el exón 4 contiene más del 90% de la región que codifica para la proteína madura.

En cabras destaca el extenso polimorfismo encontrado en el gen de la *αs1*-Cn. Además, las variantes encontradas han sido asociadas con diferencias en el contenido proteico de la leche, en la relación entre caseína y proteína total y en el rendimiento en la producción de queso (GROSCLAUDE *et al.*, 1994). Debido a la importancia de estos caracteres, se ha incluido el genotipado para la *αs1*-Cn en la evaluación de machos utilizados en inseminación artificial en Francia (BARBIERI *et al.*, 1995). También se ha encontrado polimorfismo en otras caseínas caprinas: tres variantes para el gen de la *αs2*-Cn (A, B, y C) y otras tres para el gen de la β-Cn (A, B, y O), siendo O un alelo nulo de la β-Cn. Por el contrario, no se han descrito variantes para las proteínas del lactosuero, aunque se han descrito diferentes formas que difieren en su movilidad electroforética (revisado en MOIOLI *et al.*, 1998). Recientemente, se han descrito polimorfismos en la región 3' no codificante y en la región promotora del gen de la β-lactoglobulina (PENA *et al.*, 2000; YAHYAOUI *et al.*, 2000).

Debido a la importancia de la κ-Cn en las propiedades tecnológicas de la leche, varios grupos han analizado esta proteína en cabras en busca de variantes genéticas, describiéndose la existencia de polimorfismos (DI LUCCIA *et al.*, 1990; JAUBERT y MARTIN, 1992; LAW y TZIBOULA, 1993; ANGULO *et al.*, 1994; RECIO *et al.*, 1997). Sin embargo, no se han caracterizado las diferencias en la secuencia de aminoácidos de las variantes encontradas. También, se ha descrito un polimorfismo de restricción (DI GREGORIO *et al.*, 1989) en el gen de la κ-Cn por digestión con BamHI, EcoRV y PvuII, e hibridación con una sonda correspondiente al cDNA de la κ-Cn bovina.

En este trabajo hemos evaluado la técnica Base Excision Sequence Scanning (BESS, HAWKINS y HOFFMAN, 1997) para la detección de mutaciones. Esta técnica consiste básicamente en la incorporación por PCR de cantidades limitadas de dUTP y en la posterior digestión del producto con los enzimas: Uracilo N-glicosilasa y Endonucleasa IV. La comparación de los patrones de bandas obtenidos en relación con una muestra control permite la detección de mutaciones en las que una timina está implicada.

El objetivo del presente trabajo fue el estudio del polimorfismo genético en el gen de la κ-Cn caprina. Se caracterizan, por primera vez, variantes del gen de la κ-Cn caprina, localizadas en el exón 4. Se desarrollan protocolos para el genotipado de las posiciones polimórficas y se estudian las frecuencias alélicas en diferentes razas caprinas.

Material y métodos

Muestras de animales

Un total de 147 muestras de ADN caprino han sido utilizadas en este estudio, incluyendo las razas Canaria, Malagueña, Murciano-Granadina, Payoya y Saanen. También se ha analizado una muestra de la cabra salvaje española (*Capra pyreneica sp. hispanica*). El ADN fue aislado a partir de muestras de sangre siguiendo protocolos estándar (AUSUBEL *et al.*, 1987).

Amplificación del gen de la κ-Cn caprina

Un fragmento de 459 pb del exón 4 de la k-CN caprina fue amplificado por PCR a partir de muestras de ADN genómico utilizando los *primers* Kb1 y Kb2 (figura 1), diseñados a partir de la secuencia del ADNc de la κ-Cn caprina (COLL *et al.*, 1993). Los *primers* Kb1 y Kb2 fueron marcados fluorescentemente en el extremo 5' con Hex y 6-Fam, respectivamente.

Detección de polimorfismo mediante el método BESS

Diez muestras de ADN de diferentes animales por cada raza fueron analizadas utilizando el método BESS. La reacción de PCR fue realizada con 200 μM *BESS T-Scan dNTP Mix*. Para la reacción de excisión, se utilizaron 5 μl del producto de PCR, 0,5 μl *BESS T-Scan Excision Enzyme Mix* y 0,6 μl del *BESS T-SCAN EXCISION ENZYME BUFFER*. La reacción fue incubada durante 20 min. a 37°C y enfriada en hielo. Se añadieron 25 μl de formamida y 1 μl del marcador de tamaños *GeneScan* 500 ROX (Perkin Elmer Applied Biosystems) y se desnaturalizó la muestra a 95°C durante 3 min. El producto obtenido fue analizado por electroforesis capilar y detección fluorescente utilizando un equipo de secuenciación automatizada (Genetic Analyzer 310, Perkin Elmer) y el software *GeneScan*. Gracias al marcaje de los dos *primers* con fluorescencias diferentes, se analizaron las dos cadenas del ADN simultáneamente.

Reacciones de secuenciación

Se secuenciaron directamente los productos de PCR para obtener una primera información de la secuencia nucleotídica. En las muestras donde se había encontrado polimorfismo, los productos de PCR se clonaron en plásmido (pTZ18U y pTZ19U), se aislaron clones para los diferentes alelos y se secuenciaron. Ambas cadenas del ADN fueron secuenciadas mediante el método dideoxi, utilizando un secuenciador automático de ADN (ABI Prism 310, Perkin Elmer). Varios clones para cada alelo fueron secuenciados para eliminar posibles errores en la reacción de PCR.

Genotipado mediante PCR-RFLP

Diez μl del producto de PCR fueron digeridos con 10 U del enzima de restricción *Alw*44I (Roche Molecular Diagnosys) a 37°C toda la noche y con *Bse*NI (Roche Molecular Diagnosys) a 65°C durante 6 horas. Los fragmentos obtenidos después de la digestión fueron separados por electroforésis en geles de agarosa al 2%.

Resultados

Detección de polimorfismo utilizando el método BESS

Un fragmento de 459 pb del exón 4 de la κ-Cn fue amplificado por PCR y analizado por el método BESS utilizando electroforesis capilar y detección fluorescente. Tres posibles posiciones polimórficas fueron detectadas, de las cuales 2 fueron confirmadas por secuenciación en las posiciones 471 y 509 de la secuencia del ADNc (figura 1). La tercera mutación detectada fue considerada un artefacto debido a que se localizaba en una región muy próxima al *primer* fluorescente, donde la señal y el ruido de fondo son muy elevados.

Análisis de la secuencia nucleotídica del exón 4 de la κ-Cn

Los productos de PCR de diferentes animales y razas fueron secuenciados. Se confirmaron las mutaciones previamente encontradas con el método BESS en las posiciones 471 y 509 y se identificaron 5 mutaciones adicionales en las posiciones 245, 284, 309, 583 y 591 (figura 1).

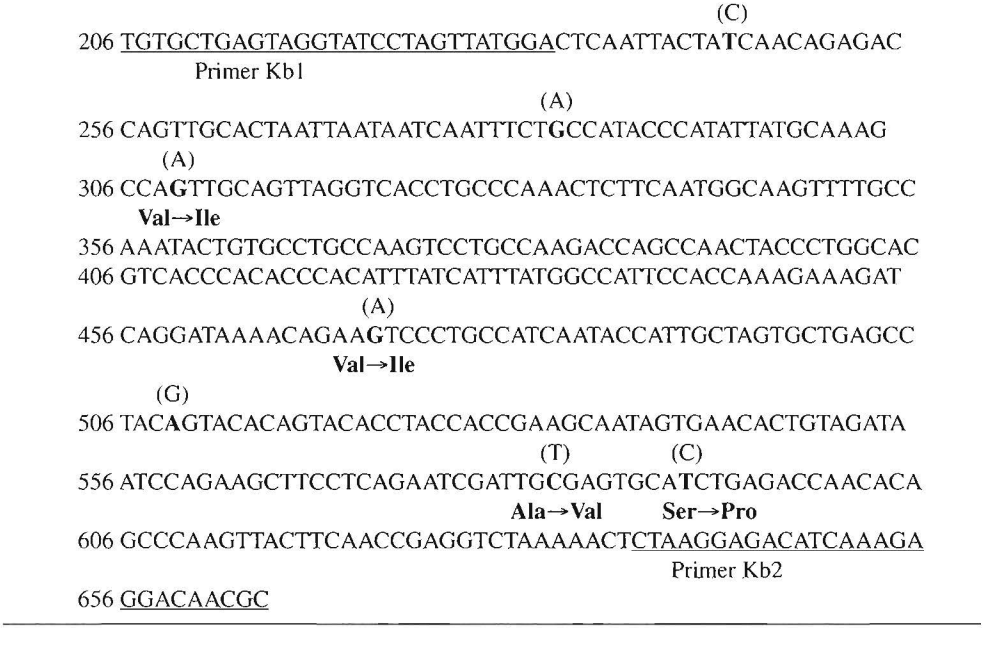


Figura 1. Secuencia nucleotídica del exón 4 de la k–Cn caprina. La numeración en el margen izquierdo corresponde a la secuencia del ADNc de la k–Cn caprina (GeneBank X60763). El doble subrayado indica la localización de los primers Kb1 y Kb2. En negritas se resaltan las posiciones polimórficas. Los nucleótidos alternativos en cada posición polimórfica se indican entre paréntesis. Las substituciones aminoacídicas se indican debajo de la secuencia

Un total de 7 posiciones polimórficas fueron detectadas en la región amplificada (459 pb) del exón 4 de la κ-Cn (figura 1). Todas corresponden a mutaciones puntuales y concretamente a transversiones en las que una T es substituida por una C o bien una G es substituida por una A. Cuatro posiciones polimórficas determinan un cambio aminoacídico en la proteína: de valina a isoleucina (en las posiciones 309 y 471), de alanina a valina (583), y de serina a prolina (591) (figura 1). Las otras tres mutaciones son silenciosas.

Caracterización molecular de nuevas variantes para el gen de la κ-Cn

Los productos de PCR correspondientes a animales de diferentes razas fueron clonados y secuenciados para establecer los haplotipos de las diferentes mutaciones detectadas. Dos clones independientes de cada animal fueron secuenciados.

En el cuadro 1 se muestran las variantes de la κ-Cn y los cambios aminoacídicos producidos. Las variantes identificadas se denominan κ-Cn A, B y C. La κ-Cn A corresponde a la secuencia nucleotídica del ADNc de la κ-Cn previamente descrita (Coll *et al.*, 1993). También se secuenció el exón 4 de la κ-Cn de la cabra salvaje, encontrándose que es diferente a la secuencia nucleotídica de los tres alelos caprinos.

Las variantes A y B difieren sólo en un aminoácido en la posición 119 (A: Valina; B: Isoleucina). La κ-Cn A difiere de la variante C en cuatro substituciones aminoacídicas (Posiciones 65 y 119: Valina/Isoleucina; Posición 156: Alanina/Valina; Posición 159: Serina/Prolina). La primera substitución aminoacídica (posición 65) se localiza en la región N-terminal de la proteína (para-κ-Cn), mientras las otras tres substituciones se localizan en la región C-terminal (caseinomacropéptido).

Genotipado por PCR-RFLP

Se han desarrollado protocolos PCR-RFLP para la detección de las mutaciones localizadas en las posiciones 309 y 591, ambas producen un cambio de aminoácido. El polimorfismo en la posición 309 (G/A) afecta a la diana de restricción *Bse*NI, que se encuentra en las variantes A y B, pero no en la variante C. La digestión del fragmento de 459 pb del exón 4 de la κ-Cn produce tres fragmentos de 54, 51, y 354 pb en los alelos A y B, mientras que el alelo C genera sólo dos fragmentos de 54 y 405 pb. La substitución nucleotídica en la posición 591 (T/C) determina la aparición de una diana de restricción *Alw*44I en el alelo C. Por tanto, la digestión del alelo C con esta endonucleasa produce dos fragmentos (78 y 381 pb) mientras que los productos de PCR de los alelos A y B no son digeridos.

El genotipado de 147 animales de diferentes razas Francesas y Españolas utilizando los dos métodos PCR-RFLP, indicaron que las dos posiciones polimórficas se encuentran siempre ligadas: G/A en la posición 309 con T/C en la posición 591. Por tanto, las variantes A y B pueden distinguirse de la variante C utilizando cualquiera de los dos métodos de genotipado por PCR-RFLP. En el cuadro 2 se muestran las frecuencias génicas obtenidas para estas variantes en las razas caprinas estudiadas.

Discusión

Un fragmento de 459 pb del exón 4 que contiene gran parte de la región codificante

de la κ–Cn caprina (141 aminoácidos de un total de 171), fue amplificado por PCR y se realizó un análisis por el método BESS y por secuenciación para la búsqueda de variantes genéticas. Se identificaron 7 posiciones polimórficas, de los cuales 2 fueron detectadas por el método BESS y 5 por secuenciación.

Debido a que todas estas mutaciones implican un cambio de una timina en una de las cadenas de ADN, deberían ser detectadas por el método BESS. Sin embargo, sólo se detectaron 2 de las 7 mutaciones y no se encontraron diferencias en el patrón de picos generados con el método BESS en las otras 3 mutaciones. Este resultado puede ser explicado por una actividad no específica del enzima de excisión del sistema BESS, cuya actividad depende de la secuencia nucleotídica en la que se encuentra la posición polimórfica. Por tanto, la técnica parece tener limitaciones de sensibilidad en la detección de mutaciones, principalmente en los extremos de la región analizada. Además, las mutaciones identificadas fueron todas detectadas en la cadena marcada con 6-Fam y no en la cadena marcada con Hex, a pesar de que ambas fluorescencias dieron picos de similar intensidad.

De las 7 posiciones polimórficas, 3 corresponden a substituciones silenciosas y las otras 4 determinan cambios aminoacídicos (figura 1). Estas 7 mutaciones son las que diferencian las tres variantes genéticas de la κ–Cn caprina encontradas, denominadas κ–Cn A, B y C (cuadro 1). Las variantes A y B difieren sólo en una substitución aminoacídica en la posición 119, mientras que la variante C difiere de la A en cuatro substituciones aminoacídicas. Las variantes A y B pueden distinguirse de la variante C utilizando cualquiera de los dos protocolos PCR-RFLP descritos anteriormente.

A pesar de que diferentes estudios habían descrito la existencia de polimorfismo en la κ–Cn caprina, no se había caracterizado ninguna variante. MERCIER *et al.* (1976a) secuenció el caseinomacropéptido a partir de la leche de una cabra Alpina-Saanen y describió la presencia en la posición 119 de dos posibles aminoácidos: valina o isoleucina. Estos autores sugirieron que el animal secuenciado era heterocigoto, pero postularon que la isoleucina era el aminoácido más probable en esta posición para la κ–Cn caprina debido a que se encontraba en otras especies relacionadas. Por el contrario, en la secuenciación del ADNc de la κ–Cn caprina (COLL *et al.*, 1993) se encontró el codón que codifica para la valina en esta posición. Esta secuencia de aminoácidos corresponde a la κ–Cn A, mientras que la substitución de valina por isoleucina en la posición 119 corresponde a la variante B.

Otros trabajos describieron la presencia de variantes de la κ–Cn que se caracterizaban probablemente por uno o más cambios aminoacídicos en la región N-terminal de la proteína: Di LUCCIA *et al.* (1990) en una raza local Italiana y Law y Tziboula (1993) en una población Británica de la raza Saanen y en una población comercial Griega. La variante encontrada corresponde a la substitución de una arginina por un residuo neutro (Di LUCCIA *et al.*, 1990). Esta variante no se correspondería a ninguno de los tres alelos descritos en el presente trabajo, ya que sólo la variante C sufre un cambio aminoacídico en la región N-terminal, pero éste no tiene efecto sobre la carga neta de la proteína.

La variante C se ha encontrado con frecuencias bajas en las razas españolas Murciano-Granadina y Canaria, pero no se ha encontrado en las razas Malagueña y Payoya. Sin embargo, el pequeño número de animales analizado no permite afirmar que esté ausente en estas razas. La frecuencia del alelo C en la raza francesa Saanen es más elevada (10%) (cuadro 1). La selección artificial para la producción de leche podría

haber influido en la distribución de estas frecuencias de la κ–Cn, siendo la Saanen francesa una raza muy seleccionada. En este sentido, el análisis del patrón de substituciones nucleotídicas entre secuencias del gen de la κ–Cn de diferentes especies de la familia *Bovidae* (que incluye la cabra), sugiere que una selección positiva ha acelerado la divergencia de la región codificante del gen (WARD *et al.*, 1997). Sin embargo, los efectos de la selección positiva en la distribución de los alelos de la κ–Cn tendría que ser evaluada en un estudio más amplio. Por el contrario, la distribución de alelos de la *αs1*-Cn caprina, estudiada en un número relativamente grande de animales de diferentes razas, no parece estar afectada significativamente por la selección para la producción de leche (JORDANA *et al.*, 1996). Probablemente, la deriva genética y el efecto fundador han jugado un papel importante en la distribución de las frecuencias alélicas de los genes de las proteínas lácteas.

Los aminoácidos de las posiciones 119 (isoleucina) y 159 (prolina) de la κ–Cn están bien conservados entre las diferentes especies analizadas, entre las que se incluyen la vaca, la oveja, el cerdo, la rata, el ratón y el hombre. Las secuencias nucleotídicas de la κ–Cn de los diferentes alelos caprinos y de la cabra salvaje española fueron alineados con las secuencias del exón 4 de otras especies relacionadas. A partir de esta alineación, se obtuvo una secuencia consensus para las 7 posiciones polimórficas caprinas. Esta secuencia es idéntica a la encontrada en la cabra salvaje española, sugiriendo que esta secuencia corresponde a la forma ancestral de la variantes caprinas actuales. El alelo B caprino es el que muestra una mayor similitud nucleotídica con la secuencia ancestral (5 de 7 posiciones nucleotídicas), mientras que el alelo C es el más divergente. Este resultado concuerda con la baja frecuencia del alelo C en las diferentes razas caprinas analizadas.

Diversos trabajos han demostrado los efectos de determinadas variantes genéticas de los genes de proteínas lácteas en la composición y propiedades físico-químicas de

Cuadro 1

Variantes del gen de la κ-Cn caprina. Se indica la posición de los nucleótidos polimorficos y sus correspondientes aminoácidos. En cada variante de la κ-Cn caprina y en la cabra salvaje se indican los nucleótidos que se encuentran en cada posición polimórfica y los cambios aminoacídicos se indican entre paréntesis

Posición nucleóotidos	Posición aminoácidos	Variantes en Cabra			Cabra salvaje	Genotipado
		A	B	C		
245	43	T	T	C	C	
284	56	G	G	A	G	
309	65	G (Val)	G (Val)	A (Ile)	G (Val)	*Bse*NI
471	119	G (Val)	A (Ile)	A (Ile)	A (Ile)	
509	131	A	A	G	A	
583	156	C (Ala)	C (Ala)	T(Val)	C (Ala)	
591	159	T (Ser)	T (Ser)	C (Pro)	C (Pro)	*Alw*44I

Cuadro 2
Distribución de las frecuencias génicas del gen de la κ-Cn caprina en razas Españolas y Francesas. Las frecuencias génicas se obtuvieron genotipando los animales por PCR-RFLP con BseNI y Alw44I. Los alelos A y B tienen una guanina y una timina en las posiciones 309 y 591, respectivamente. El alelo C tiene adenosina y citosina en estas mismas posiciones, respectivamente

Raza	Número de animales	Frecuencias génicas	
		A+B	C
Malagueña	17	1	0
Payoya	11	1	0
Canaria	48	0,989	0,011
Murciano-Granadina	38	0,986	0,014
Saanen	33	0,894	0,106

la leche. En bovino, la variante B de la κ–Cn ha sido asociada con un contenido más elevado de caseína total en la leche y con una relación caseínas / proteínas séricas mayor (revisado en KAMINSKI *et al.*, 1996). Estas propiedades son importantes principalmente en el rendimiento de la producción de queso. El genotipo de la κ–Cn se incluye en algunos catálogos de inseminación artificial y, además, la selección a favor del alelo B se practica en algunos países (NG-KWAI HANG, 1998).

En caprino, los diferentes alelos de la *αs1*-Cn determinan un contenido diferente de esta proteína en la leche, con cuatro niveles de expresión que varían desde cero para el alelo nulo hasta 3·6 g/l para los alelos A, B, y C (GROSCLAUDE *et al.*, 1994). Serán necesarios nuevos trabajos en diferentes razas caprinas para analizar la distribución de los alelos de la κ–Cn descritos en este trabajo, así como para evaluar su posible efecto sobre las propiedades tecnológicas de la leche. El método de genotipado por PCR-RFLP descrito en este trabajo será de utilidad en el genotipado de un mayor número de animales.

Agradecimientos

Este trabajo ha sido realizado gracias a una beca concedida a M.H.Y. del CIHEAM (Instituto Agronómico Mediterráneo de Zaragoza). Agradecemos a M. Amills y J. Capote por haber facilitado algunas de las muestras de ADN analizadas.

Bibliografía

ADDEO F., SOULIER S., PELISSIER J.P., CHOBERT J.M., MERCIER J.C., RIBADEU-DUMAS B,. 1978. Preparation and fractionation of goat kappa-casein: analysis of the glycan and peptide components. *Journal of Dairy Research* 45 191-196.

ANGULO C., DÍAZ-CARRILLO E., MUÑOZ A., ALONSO A., JIMÉNEZ I., SERRADILLA J.M., 1994. Effect of electrophoretic goat's κ–casein polymorphism on milk

yield and main components yield. *V World Congress on Genetic Applied to Livestock Production* Guelfh 19 333-336.

AUSUBEL F.M., BRENT R., KINGSTON R.E., MOORE D.D., SEIDMEN G.G., SMITH J.A., y STRUHL K., (Ed.), 1987. *Current Protocols in Molecular Biology* Green publishing Associates and Wiley-Interscience.

BARBIERI M.E., MANFREDI E., ELSEN J.M., RICORDEAU G., BOUILLON J., GROSCLAUDE F., MAHÉ M.F., BIBÉ B., 1995. [Influence of αs1 casein locus on the milk performances and genetic parameters in the Alpine goat] *Genetics Selection Evolution* 27 437-450.

COLL A., FOLCH J.M., SÁNCHEZ A., 1993. Nucleotide sequence of the goat κ-casein cDNA *Journal of Animal Science* 71 2833.

COLL A., FOLCH J. M., SÁNCHEZ A., 1995. Structural features of the 5' flanking region of the caprine κ-casein gene *Journal of Dairy Science* 78 973-977.

DI GREGORIO P., RANDO A., RAMUNNO L., MASINA P., PERAGOSTINI E., 1989 [Polymorphism of the DNA region containing casein genes in sheep and goat] XXIV *Simposio Internazionale Zootecnica* pp 275-282.

DI LUCCIA A., MAURIELLO R., CHIANESE L., MOIO L., y ADDEO F., 1990. Kappa casein polymorphism in caprine milk *Scienza e Tecnica Lattiero-Casearia* 41 305-314.

GROSCLAUDE F., RICORDEAU G., MARTIN P., REMEUF F., VASSAL L., BOUILLON J., 1994. [From gene to cheese: the polymorphism of the caprin αs1-casein. its effects and evolution] INRA *Productions Animales* 7 3-19.

HAWKINS G.A., HOFFMAN L.M., 1997. Base excision sequence scanning: a new method for rapid sequence scanning and mutation detection *Nature Biotechnology* 15 803-804.

JAUBERT A., MARTIN P., 1992. Reverse phase HPLC analysis of goat caseins. Identification of αs1 and αs2 genetic variants. Lait 72 235-247.

JORDANDA J., AMILLS M., DÍAZ E., ANGULO C., SERRADILLA J.M., SÁNCHEZ A., 1996. Gene frequencies of caprine αs1-casein polymorphism in Spanish goat breeds *Small Ruminant Research* 20 215-221.

KAMINSKI S., 1996. Bovine κ-casein gene: molecular nature and application in dairy cattle breeding *Journal of Applied Genetics* 37 179-196.

LAW A.J.R., TZIBOULA A., 1993. Fractionation of caprine κ-casein and examination of polymorphism by FPLC *Milchwissenschaft* 48 68-71.

MERCIER J.C., ADDEO F., y PELISSIER J.P., 1976a. [Primary structure of the caprine caseinomacropeptide] *Biochimie* 58 1303-1310.

MERCIER J.C., CHOBERT J.M., ADDEO F., 1976b. Comparative study of the amino acid sequences of the caseinomacropeptides from seven species *FEBS Letters* 72 208-214.

MOIOLI B., PILLA F., TRIPALDI C., 1998. Detection of milk protein genetic polymorphisms in order to improve dairy traits in sheep and goats: a review *Small Ruminant research* 27 185-189.

NG-KWAI-HANG K.F., 1998. Genetic polymorphism of milk proteins: relationships with production traits, milk composition and technological properties *Canadian Journal of Animal Science* 78 (Supplement) 131-147.

PENA R.N., SÁNCHEZ A., FOLCH J.M., 2000. Characterisation of genetic polymorphism in the goat β-lactoglobulin gene *Journal of Dairy Research* 67 217-224.

RECIO I., PÉREZ-RODRÍGUEZ M.L., AMIGO L., RAMOS M., 1997. Study of the polymorphism of caprine milk caseins by capillary electrophoresis *Journal of Dairy Research* 64 515-523.

SWAISGOOD H.E., 1992. Chemistry of the caseins. In *Advanced Dairy Chemistry-1. Proteins*, pp. 63-110 (Ed. P. F. FOX). London: Elsevier Applied Science.

WARD T.J., HONEYCUTT R.L., DERR J.N., 1997. Nucleotide sequence evolution at the kappa-casein locus: evidence for positive selection within the family Bovidae *Genetics* 147 1863-1872.

YAHYAOUI M.H., PENA R.N., SÁNCHEZ A. FOLCH J.M., 2000. Rapid communication: polymorphism in the goat b-lactoglobulin proximal promoter region *Journal of Animal Science* 78: 1100-1101.

ZITTLE C.A., CUSTER J.H., 1966. Identification of the κ-casein among the components of whole goat casein *Journal of Dairy Science* 49 788-791.

ITEA (2000), Vol. 96A N.° 3, 311-317

ADECUACIÓN DE LA ESTRUCTURA DE DATOS DEL CABALLO DE CARNE EN ÁLAVA PARA UNA ESTIMA DE COMPONENTES DE VARIANZA

A. Legarra*, E. Ugarte*, E. Ruiz de Zárate**

* NEIKER A. B. Granja Modelo Arkaute, Apdo. 46, 01080 Vitoria-Gasteiz, España

** SERGAL, La Paloma, n.° 4, bajo, 01002 Vitoria-Gasteiz, España

RESUMEN

Se han simulado diferentes poblaciones de acuerdo a la estructura de los datos genealógicos y productivos existente en la asociación de ganaderos del caballo de carne de Álava para determinar si dicha estructura permite una correcta estima de componentes de varianza. El modelo utilizado para simular la población fue un modelo aditivo con efectos directos y maternos y efecto del grupo de comparación Rebaño-año. Se simularon un total de 27 poblaciones para las diferentes combinaciones de la heredabilidad del efecto directo (h_d^2 = 0,1, 0,25 y 0,40), la heredabilidad del efecto materno (h_m^2 = 0,1, 0,25 y 0,40) y la correlación genética entre ambos componentes (r_{dm} = 0,2, -0,2 y -0,40). De cada población se hizo un total de 100 repeticiones.

Posteriormente se estimaron los componentes de varianza de los datos simulados mediante una estima REML para un modelo que contemplaba efectos directos y maternos y para otro modelo que sólo contemplaba efectos directos. Tanto las estimas de la heredabilidad del efecto aditivo directo como del materno fueron precisas, sin embargo la estima de la correlación genética entre ambas fue muy imprecisa. Las estimas obtenidas con el modelo que tenía en cuenta sólo el efecto directo sobreestimaron la heredabilidad de este último.

Se concluye que la estructura de datos existente permitiría estimar aceptablemente los componentes de varianza, siempre y cuando se trabaje con un modelo que incluya efectos maternos.

Palabras clave: Componentes de varianza, Estructura de datos, Efectos maternos, Caballo de carne.

SUMMARY

Different populations were simulated according to the genealogical and productive structure of Alava meat horse breeders' association in order to determine the reliability of a variance component estimation in the real population. The simulation model consisted of an additive model, which included herd-year as fixed effect and direct and maternal genetic additive effects. A total of 27 populations were simulated for every combination of the following genetic parameters: heritability of direct effect (h_d^2 = 0.1, 0.25 and 0.40), heritability of maternal effect (h_m^2 = 0.1, 0.25 y 0.40), and genetic correlation between both effects (r_{dm} = 0.2, -0.2 and -0.40). Each population was simulated 100 times.

Afterwards, variance components were estimated from the simulated data using a REML method. In this analysis, two models were considered: the first one dealing with direct and maternal effects, and the second one including only direct effects. In the first model, estimates of direct and maternal effects were accurate, but estimates of genetic correlation showed lack of accuracy. In the second one, the obtained estimates resulted in a highly positive bias.

It can be concluded that the data structure could allow an acceptable variance component estimation, provided that maternal effects are included in the model.

Key words: Variance components, Data structure, Maternal effects, Meat horse.

Introducción

En 1984 se creó en Alava la asociación de ganaderos de equino de Álava (ASGAEQUINO). Forman parte de ella 85 explotaciones que suponen el 25% de las explotaciones existentes en Álava y recogen el 45% del censo existente en este territorio histórico. El 95% de estas cabezas corresponde a caballos de tipo pesado o semipesado que actualmente se destinan a la producción de carne de potro (RUIZ DE ZÁRATE, 1999). Desde su formación, la asociación ha llevado a cabo varios programas entre los que cabe destacar el programa de identificación individual, el programa sanitario, el programa reproductivo y el programa de control de rendimientos. En 1997 se realizó asimismo un estudio de caracterización morfológica que ha servido de base para determinar el estándar racial de los animales y ha permitido la tramitación de su libro genealógico (BOPV, 16/8/99) y que esta agrupación de animales fuera admitida como raza en el catálogo de razas autóctonas españolas como "Caballo de Monte del País Vasco-Euskal Herriko Mendiko Zaldia" (BOE, 21/3/2000). Actualmente, dentro de un proyecto de investigación financiado por el Gobierno Vasco se está trabajando en la caracterización del sistema de producción del caballo de carne y dentro de los objetivos específicos del proyecto está la identificación de los criterios y objetivos para el diseño de un programa de mejora para esta población. Aunque la actividad económica que lleva asociada es pequeña (el consumo de carne de potro en España y Álava supone respectivamente el 0,2% y 0,5% del consumo de carne total; FAO, 1999; DAP, 1997) el apoyo que desde la administración se da a esta especie se justifica al tener en cuenta que la Comunidad Autónoma del País Vasco (CAPV) y Navarra son las únicas zonas dentro de la Unión Europea donde el sistema de producción del caballo de carne es totalmente extensivo maximizando el aprovechamiento del medio natural y ayudando al mantenimiento de la biodiversidad y al mantenimiento de la población en zonas rurales. Ello reporta a la sociedad una serie de valores añadidos de tipo social y económico.

El objetivo que buscan los ganaderos es el de obtener potros que presenten al destete buenos crecimientos a pesar de estar criados en base a un sistema de producción extensivo. El disponer de datos de pesos al nacimiento y de pesos al destete nos permite calcular ese dato. Durante este período el potro es amamantado por su madre, por lo que a la hora de analizar genéticamente esos datos

parece lógico considerar en el modelo los efectos aditivos directo y materno.

En este trabajo se presentan los resultados de uno de los análisis que se han realizado para determinar si la estructura poblacional existente en la actualidad podría permitir trabajar de forma adecuada en el contexto de un programa de mejora genética a la hora de realizar estimas adecuadas de los parámetros genéticos, o bien sería necesario modificar dicha estructura.

Materiales y métodos

Para hacer el estudio se han simulado diferentes poblaciones de acuerdo a los parámetros y estructura de los datos genealógicos y productivos existente en ASGAEQUINO. Esta base de datos recoge información sobre 5509 partos de los cuales tenemos medidas de crecimiento diario. Estos partos corresponden a un total de 2151 yeguas que están repartidas en 106 explotaciones durante 16 años. El fichero de genealogías relacionadas recoge un total de 7570 animales en el pedigrí. El cuadro 1 muestra las características de la población simulada.

Ante la falta de estudios previos correspondientes a estimas de estos componentes en caballo de carne se han tomado como referencia valores obtenidos para el mismo carácter en vacuno de carne (QUINTANILLA y PIEDRAFITA, 2000; BENNETT y GREGORY, 1996; GUTIÉRREZ *et al.*, 1997). Se simularon un total de 27 poblaciones para las diferentes combinaciones de la heredabilidad del efecto directo (h_d^2 = 0,1, 0,25 y 0,40), la heredabilidad del efecto materno (h_m^2 = 0,1, 0,25 y 0,40) y la correlación genética entre ambos componentes (r_{dm}= 0,2, -0,2 y -

Cuadro 1
Características de la población simulada

Parámetro	Valor
N.º de animales en datos (todos con madre conocida)	5.509
N.º de animales en datos con padre conocido	4.822
N.º de padres de animales en datos	225
N.º de madres de animales en datos	2.151
N.º de padres con su propio dato de crecimiento	135
N.º de madres con su propio dato de crecimiento	879
N.º medio de hijos en datos /madre	2,56
N.º medio de hijos en datos /padre	21,40
N.º de animales diferentes en genealogía	7.570
N.º de padres diferentes en genealogía	235
N.º de madres diferentes en genealogía	2.214
N.º de rebaños	106
N.º medio de padres diferentes en cada rebaño-año	1,42
N.º de rebaños con un solo padre en datos	17
N.º de grupos de comparación rebaño-año	797
N.º medio de padres diferentes en cada rebaño	4,07
N.º de rebaños-año con un solo padre en datos	500

0,40). De cada población se hizo un total de 100 repeticiones.

El modelo utilizado para simular la población fue un modelo aditivo que incluía los efectos directo y materno y efecto del grupo de comparación Rebaño-año como efecto fijo, de acuerdo al siguiente modelo:

$$y = Xb + Z_d u_d + Z_m u_m + e$$

donde

y es el vector de observaciones.

b es el vector de efectos fijos (Rebaño-año).

u_d y u_m son los vectores de efectos genéticos directos y maternos.

X, Z_d y Z_m son las matrices de incidencia que relacionan el vector de observaciones con los efectos correspondientes.

En esta simulación sólo se tomó como efecto fijo el efecto grupo de comparación para tener en cuenta su pequeño tamaño y los riesgos de desconexión. Este efecto del grupo de comparación se muestreó de una distribución uniforme acotada de una amplitud de 5 veces la desviación estándar del carácter, de acuerdo a estudios previos realizados sobre datos reales.

Las distribuciones de muestreo de los efectos aleatorios fueron las siguientes:

$$(u_d\ u_m) \sim N((0\ 0), G_0 \otimes A)$$

$$e \sim N(0, I\ \sigma_e^2)$$

siendo A la matriz de correlaciones genéticas aditivas entre los individuos.

G_0 la matriz de varianzas y covarianzas de los efectos genéticos aditivos directo y materno.

I la matriz identidad.

σ_e^2 la varianza residual.

Para realizar la estima de componentes de varianza de los datos simulados se utilizó el programa VCE 4.2.5 por un método REML de gradientes analíticos (NEUMAIER y GROENEVELD, 1998), según un modelo con efectos maternos idéntico al que se utiliza en la simulación. Para completar el estudio se realizó el mismo análisis sin contemplar el efecto materno.

Cuadro 2
Valores obtenidos para la estima de la heredabilidad del efecto aditivo directo, $\hat{h}_d^2$ (media ± desv. est.), en cada combinación simulada de componentes de varianza

		$h_d^2 = 0,1$	$h_d^2 = 0,25$	$h_d^2 = 0,40$
$h_m^2 = 0,1$	$r_{dm} = +0,20$	0,10 ± 0,05	0,24 ± 0,06	0,41 ± 0,07
	$r_{dm} = -0,20$	0,12 ± 0,05	0,26 ± 0,07	0,40 ± 0,07
	$r_{dm} = -0,50$	0,11 ± 0,05	0,26 ± 0,07	0,39 ± 0,07
$h_m^2 = 0,25$	$r_{dm} = +0,20$	0,10 ± 0,05	0,25 ± 0,06	0,39 ± 0,07
	$r_{dm} = -0,20$	0,11 ± 0,05	0,25 ± 0,07	0,39 ± 0,07
	$r_{dm} = -0,50$	0,11 ± 0,05	0,24 ± 0,06	0,39 ± 0,06
$h_m^2 = 0,40$	$r_{dm} = +0,20$	0,10 ± 0,04	0,26 ± 0,05	0,39 ± 0,07
	$r_{dm} = -0,20$	0,10 ± 0,05	0,24 ± 0,07	0,40 ± 0,06
	$r_{dm} = -0,50$	0,11 ± 0,05	0,25 ± 0,06	0,40 ± 0,07

Cuadro 3
Valores obtenidos para la estima de la heredabilidad del efecto aditivo materno, $\hat{h}_m^2$ (media ± desv. est.), en cada combinación simulada de componentes de varianza

		$h_d^2 = 0,1$	$h_d^2 = 0,25$	$h_d^2 = 0,40$
$h_m^2 = 0,1$	$r_{dm} = + 0,20$	0,12 ± 0,04	0,11 ± 0,04	0,10 ± 0,04
	$r_{dm} = - 0,20$	0,11 ± 0,04	0,10 ± 0,04	0,10 ± 0,04
	$r_{dm} = - 0,50$	0,10 ± 0,04	0,10 ± 0,04	0,10 ± 0,04
$h_m^2 = 0,25$	$r_{dm} = + 0,20$	0,25 ± 0,04	0,25 ± 0,04	0,25 ± 0,04
	$r_{dm} = - 0,20$	0,24 ± 0,04	0,25 ± 0,04	0,25 ± 0,04
	$r_{dm} = - 0,50$	0,25 ± 0,04	0,25 ± 0,04	0,25 ± 0,04
$h_m^2 = 0,40$	$r_{dm} = + 0,20$	0,40 ± 0,03	0,39 ± 0,03	0,40 ± 0,03
	$r_{dm} = - 0,20$	0,39 ± 0,03	0,40 ± 0,03	0,40 ± 0,03
	$r_{dm} = - 0,50$	0,39 ± 0,03	0,40 ± 0,03	0,40 ± 0,03

Cuadro 4
Valores obtenidos para la estima de la correlación entre los efectos aditivos directo y materno, $\hat{r}_{dm}$ (media ± desv. est.), en cada combinación simulada de componentes de varianza

		$h_d^2 = 0,1$	$h_d^2 = 0,25$	$h_d^2 = 0,40$
$h_m^2 = 0,1$	$r_{dm} = + 0,20$	0,7 ± 0,38	0,22 ± 0,33	0,24 ± 0,28
	$r_{dm} = - 0,20$	-0,24 ± 0,38	-0,17 ± 0,31	-0,17 ± 0,24
	$r_{dm} = - 0,50$	-0,49± 0,34	-0,46 ± 0,24	-0,48 ± 0,21
$h_m^2 = 0,25$	$r_{dm} = + 0,20$	0,22 ± 0,33	0,26 ± 0,25	0,26 ± 0,24
	$r_{dm} = - 0,20$	-0,15 ± 0,32	-0,17 ± 0,22	-0,18 ± 0,16
	$r_{dm} = - 0,50$	-0,49 ± 0,27	-0,49 ± 0,16	-0,48 ± 0,15
$h_m^2 = 0,40$	$r_{dm} = + 0,20$	0,20 ± 0,26	0,22 ± 0,19	0,21 ± 0,16
	$r_{dm} = - 0,20$	-0,17 ± 0,28	-0,17 ± 0,17	-0,17 ± 0,13
	$r_{dm} = - 0,50$	-0,50 ± 0,21	-0,50 ± 0,11	-0,50 ± 0,09

Resultados

En los cuadros 2, 3 y 4 se pueden ver las estimas REML que se obtuvieron mediante el VCE para cada uno de los ratios de componentes de la varianza ($\hat{h}_d^2$, $\hat{h}_m^2$ y $\hat{r}_{dm}$ respectivamente) en las poblaciones simuladas.

Se observa que las heredabilidades de los efectos aditivos directo (cuadro 2) y materno (cuadro 3) se estiman prácticamente sin sesgo (respecto al valor simulado), y con un

pequeño error de muestreo en todos los casos considerados. Sin embargo, la correlación genética aditiva (cuadro 4) no se estima tan bien, de tal forma que presenta altos sesgos y desviaciones estándar; se observa que su estima es más correcta cuanto más importancia tienen los efectos genéticos aditivos en los datos finales (es decir, en aquellas poblaciones que presentan altas heredabilidades).

Además, en el cuadro 5 se observa que en algunas de las estimas de componentes de varianza se obtuvieron valores de +1 (o -1) en la correlación $\hat{r}_{dm}$.

Estos valores de +1 o –1 podrían ser producto de un error de cálculo en la estima, determinado en parte por la estructura de los datos, tanto más cuanto que son picos aislados en los histogramas de la distribución de $\hat{r}_{dm}$. Se observa que desaparecen al aumentar las heredabilidades materna y directa. Estas estimas se consideraron incorrectas y no se incluyeron en los datos de las tablas 2, 3 y 4.

En el cuadro 6 se muestra el resultado de estimar el componente de efecto directo ($\hat{h}_d^2$,), sin incluir en el modelo de análisis el efecto materno.

Como era predecible, la heredabilidad del efecto directo se estima muy mal, obteniéndose sobreestimas en todos los casos, aunque la misma sea mucho mayor de lo que en un principio se esperaría. Es posible que esto se deba a la peculiar estructura de datos existente.

Conclusiones

1. Consideramos que la estructura actual de datos permite una estima de componentes de varianza con un error aceptable.

2. El efecto materno debe incluirse obligatoriamente en el análisis genético que se realice; no hacerlo llevaría a sobreestimar de forma importante la heredabilidad.

Cuadro 5

Número de veces en las que se estima de la correlación entre los efectos aditivos directo y materno, $\hat{h}_m^2$ (media ± desv. est.), en cada combinación simulada de componentes de varianza

		$\hat{h}_d^2 = 0,1$		$h_d^2 = 0,25$		$h_d^2 = 0,40$	
		$\hat{r}_{dm} = -1$	$\hat{r}_{dm} = +1$	$\hat{r}_{dm} = -1$	$\hat{r}_{dm} = +1$	$\hat{r}_{dm} = -1$	$\hat{r}_{dm} = +1$
$h_m^2 = 0,1$	$r_{dm} = + 0,20$		19		14		9
	$r_{dm} = - 0,20$	3	7		5		1
	$r_{dm} = - 0,50$	5	10		3		
$h_m^2 = 0,25$	$r_{dm} = + 0,20$		8		2		2
	$r_{dm} = - 0,20$		1				
	$r_{dm} = - 0,50$	7					
$h_m^2 = 0,40$	$r_{dm} = + 0,20$		3				
	$r_{dm} = - 0,20$		2				
	$r_{dm} = - 0,50$	6					

Cuadro 6

Valores obtenidos sin incluir el efecto materno en el modelo para la estima de la heredabilidad del efecto aditivo directo, $\hat{h}_d^2$ (media ± desv. est. de 100 repeticiones), en cada combinación simulada de componentes de varianza

		$h_d^2 = 0,1$	$h_d^2 = 0,25$	$h_d^2 = 0,40$
$h_m^2 = 0,1$	$r_{dm} = + 0,20$	0,37 ± 0,04	0,54 ± 0,04	0,68 ± 0,04
	$r_{dm} = - 0,20$	0,28 ± 0,05	0,40 ± 0,04	0,54 ± 0,04
	$r_{dm} = - 0,50$	0,20 ± 0,05	0,29 ± 0,04	0,41 ± 0,05
$h_m^2 = 0,25$	$r_{dm} = + 0,20$	0,67 ± 0,03	0,80 ± 0,03	0,90 ± 0,03
	$r_{dm} = - 0,20$	0,57 ± 0,04	0,68 ± 0,04	0,78 ± 0,04
	$r_{dm} = - 0,50$	0,50 ± 0,04	0,54 ± 0,04	0,63 ± 0,04
$h_m^2 = 0,40$	$r_{dm} = + 0,20$	0,86 ± 0,03	0,94 ± 0,03	0,98 ± 0,01*
	$r_{dm} = - 0,20$	0,78 ± 0,03	0,86 ± 0,03	0,93 ± 0,03
	$r_{dm} = - 0,50$	0,73 ± 0,03	0,76 ± 0,03	0,81 ± 0,03

*Se desecharon 58 estimas con $\hat{\sigma}_e^2 = 0$.

Agradecimientos

Este trabajo ha sido realizado gracias a la ayuda concedida por el Gobierno vasco y a una beca concedida por el INIA a A. Legarra.

Bibliografía

BENNETT G.L., GREGORY K.E., 1996. Genetic (Co)variances among Birth Weight, 200-day Weight, and Postweaning Gain in composites and parental breeds of beef cattle. J. Anim. Sci., 64, 2598-2611.

BOE, 21/3/2000 N.° 5455. Ministerio de Agricultura, Pesca y Alimentación. Orden del 3 de marzo de 2000.

BOPV, 16/8/99. N.° 3549. Departamento de Agricultura y Pesca del Gobierno Vasco. Orden del 21 de Julio de 1999.

DAP. 1999. Anuario estadístico del sector agroalimentario 1997-1998 C.A.E. Gobierno Vasco.

FAO. 1999. FAOSTAT statistical database. URL: http://apps.fao.org/default.htm.

GUTIÉRREZ J.P., CAÑÓN J., GOYACHE F., 1997. Estimation of direct and maternal genetic parameters for preweaning traits in the Asturiana de los Valles beef cattle breed through animal and sire models. J. Anim. Br. Genet., 114, 261-266.

NEUMAIER A., GROENEVELD E., 1998. Restricted Maximum Likelihood estimation of covariances in sparse linear models. Genet. Sel. Evol., 30, 3-26.

QUINTANILLA R., PIEDRAFITA J., 2000. Efectos maternos en el peso al destete del ganado vacuno de carne: una revisión. ITEA, 96A (1), 7-39.

RUIZ DE ZÁRATE E., 1999. Caracterización del sistema de producción del Caballo de Monte del País Vasco- Euskal Herriko Mendiko Zaldia. 118 pp. Proyecto fin de carrera. ETSIA. Universidad Pública de Navarra.

ITEA (2000), Vol. **96A N.° 3,** 318-336

EVALUACIÓN DE MODELOS ALTERNATIVOS PARA EL ANÁLISIS DE DATOS DE PRODUCCIÓN DE LECHE A LO LARGO DE LA LACTACIÓN

P. López-Romero, M.J. Carabaño

Departamento de Mejora Genética y Biotecnología, INIA, Carretera de La Coruña km 7, 28040 Madrid, España

RESUMEN

Se han estimado parámetros genéticos para producción de leche en primeras lactaciones de ganado frisón español a través de modelos basados en el uso directo de las producciones en el día del control. Se han utilizado distintos modelos alternativos que fueron comparados en términos de bondad de ajuste y capacidad de predicción usando validación cruzada. Los modelos evaluados fueron un modelo de repetibilidad (MMR) y distintos modelos de regresión aleatoria (MRA), que incluyeron funciones lactacionales y polinomios de Legendre como funciones lineales para definir las regresiones aleatorias. Los MRA fueron superiores al MMR. Dentro de los MRA, aquellos modelos que ajustaron polinomios de Legendre en las regresiones aleatorias fueron mejores que los que ajustaron funciones lactacionales, presentado el inconveniente de modelizar mal los parámetros genéticos a partir de órdenes de ajuste superiores a cuatro para el efecto aditivo. Este problema no se presentó en órdenes elevados de ajuste sobre la parte permanente, habiéndose observado además que incrementos de orden del polinomio sobre el efecto permanente producían mejoras más relevantes que si el aumento de orden se producía en el efecto aditivo.

Palabras clave: Parámetros genéticos, Producciones en el día del control, Modelos de regresión aleatoria, Polinomios de Legendre, Comparación de modelos, Validación cruzada.

SUMMARY

EVALUATION OF ALTERNATIVE MODELS FOR ANALYZING MILK YIELD DATA ACROSS LACTATION

Genetic parameters for milk yield of first lactation Spanish Hosltein cattle were estimated using test day models. Different alternative models were employed and were compared in terms of goodness of fit and predictive ability using cross validation techniques. Models used were repeatilbility model (MMR) and different random regression models (MRA), using Wilmink function, Ali-Schaeffer function and Legendre polynomials. The MRA were better than MMR. Within the MRA, those models which fitted Legendre polynomials were better than those which fitted lactational funtions. Moreover, the MRA had the disadvantage that they did not describe well the genetic parameters at the onset and at the end of lactation when the order of fit for additive effect was greater than four. This problem was not observed when the high orders of fit for permanent effect were used. Furthemore, when the order of fit for permanent effect increased, the improving of the model was better than when the increasing of order was carried out for the additive effect.

Key words: Genetic parameters, Test day yields, Random regression models, Legendre polynomials, Model comparison, Cross validation.

Introducción

La utilización de los datos de los controles mensuales, a partir de ahora, producciones en el día del control (PDC), para la valoración de caracteres de producción de leche en ganado vacuno presenta ventajas en relación al uso del dato resumen de producción en 305 días (ver, p.ej. REKAYA y col., 1999). En contrapartida, la modelización de este tipo de información es, potencialmente, más compleja y los costes computacionales asociados a la aplicación de estos modelos se incrementa de forma muy importante, pudiendo ser limitante en algunos casos.

La forma de modelizar las PDC puede variar entre el modelo más sencillo que considera los datos de controles sucesivos como medidas repetidas del mismo carácter (dentro y/o entre lactaciones) hasta un modelo multicarácter de orden hipotéticamente infinito o muy elevado que considerase la producción en cada momento de la lactación como un carácter diferente. Existen numerosos estudios que han estudiado la forma de modelizar estos datos en vacuno lechero. PTAK y SCHAEFFER (1993) usaron un modelo de repetibilidad simple considerando que las producciones observadas en controles sucesivos de un mismo animal como medidas repetidas de un mismo carácter. REENTS y col. (1995a y b) desarrollaron un modelo que considera los datos de controles dentro de una misma lactación como medidas repetidas y como caracteres diferentes cuando se trata de datos de distintas lactaciones, siendo este el modelo usado actualmente en las valoraciones oficiales basadas en los datos de controles individuales en Alemania.

Un refinamiento con respecto al modelo de repetibilidad simple son los modelos de regresión aleatoria (MRA) (SCHAEFFER y DEKKERS, 1994). En este tipo de modelos se reconoce la posibilidad de ajustar curvas lactacionales diferentes para cada animal, tanto en relación a efectos sistemáticos como a los efectos genéticos, reconociendo, por ejemplo, la existencia de diferencias en el potencial genético para factores como la persistencia de las curvas de lactación. Además, estos modelos reconocen la existencia de (co)varianzas que cambian con el tiempo, tanto a nivel del componente genético, como de otros componentes ambientales no sistemáticos. Un modelo de regresión aleatoria es el empleado en la evaluación oficial de Canadá (JAMROZIK y col., 1997).

Previo a la introducción de modelos de regresión aleatoria para modelizar las PDC a lo largo de la lactación, KIRKPATRICK y col. (1990) propusieron el uso de polinomios ortogonales (polinomios de Legendre) para ajustar funciones de covarianza que van a definir las (co)varianzas entre observaciones tomadas a lo largo del tiempo. Este tipo de modelos se conoce como modelos de funciones de covarianza (MFC). MEYER y HILL (1997) mostraron que un MRA puede ser también usado para modelizar las (co)varianzas entre observaciones. VAN DER WERF y col. (1999) demostraron que para cada MFC existe un MRA equivalente. De esta manera, se puede considerar como MRA tanto aquellos que ajustan funciones lactacionales a los efectos aleatorios como los que usan polinomios de Legendre para ajustar funciones de covarianza.

Paralelamente al desarrollo de MRA bajo curvas lactacionales y de MFC, WIGGANS y GODDARD (1997) desarrollan un procedimiento alternativo orientado también a reducir la elevada dimensionalidad del modelo multicarácter. Estos autores proponen el uso de la transformación canónica sobre la matriz de (co)varianzas para el posterior análisis de un número reducido de las resultantes variables canónicas. VAN DER WERF y col. (1999) indican que esta aproximación puede

interpretarse como una versión discontinua de los MFC propuestos por KIRKPATRICK y col. (1990).

Más recientemente, PLETCHER y GEYER (1999) sugieren el uso de métodos paramétricos, basados en la utilización de funciones de distribución en lugar de los métodos no paramétricos como es el de los polinomios de Legendre como funciones de covarianza. Estos autores indican que de esta forma es posible modelizar de forma más acorde con la biología del proceso las covarianzas, reduciendo también el número de parámetros, frente al uso de polinomios que son funciones totalmente arbitrarias, con un comportamiento excesivamente oscilante cuando el grado del polinomio es alto y poco acorde con la naturaleza gradual y con límites asintóticos de los procesos biológicos. Sin embargo, no es obvio cual sería la función de distribución adecuada y estos autores muestran que una mala elección de esta función puede conducir a resultados mediocres en el ajuste de las verdaderas covarianzas.

Los MRA que usan funciones lactacionales y aquellos que usan los polinomios de Legendre para ajustar las funciones de covarianza son los más utilizados en la modelización de las PDC. Además de las ventajas teóricas de estos modelos en cuanto a flexibilidad en la modelización del mérito genético y de la estructura de covarianzas, su ventaja sobre los modelos de repetibilidad simple ha quedado reflejada en varios estudios, en cuanto a una reducción sobre la varianza residual del modelo (JAMROZIK y col., 1996 y REKAYA y col., 1999). Sin embargo, no está claro cuál es la función a utilizar y tanto los modelos lactacionales como los que usan polinomios de Legendre han mostrado comportamientos anómalos en los extremos de la lactación. Por otra parte, los trabajos que han abordado la comparación de modelos alternativos (JAMROZIK y col, 1997; OLORI y col, 1999a; VAN DER WERF, 1998) han basado la misma en la utilización de tests estadísticos (principalmente tests χ^2) y en el comportamiento de las varianzas y correlaciones para distintos momentos de la lactación, tomando como referencia los valores del modelo multicarácter. No se han considerado, sin embargo, criterios como la capacidad de ajuste de los datos o la capacidad de predicción de datos futuros.

En este trabajo se presentan resultados que permiten la comparación de funciones lactacionales y polinómicas en la modelización de las PDC en nuestra población de ganado frisón, intentando abordar esta comparación desde una perspectiva más amplia que la llevada a cabo en estudios previos sobre esta y otras poblaciones de ganado vacuno lechero.

Material y métodos

Se han utilizado producciones en el día del control (PDC) de primeras lactaciones procedentes de Navarra a partir de datos suministrados por CONAFE. Se han seleccionado únicamente aquellos rebaños con registros en todos los años del intervalo comprendido entre 1988 y 1998, además, de estos rebaños, sólo fueron tenidos en cuenta los datos pertenecientes a ese periodo. El motivo de esta selección fue debido a la necesidad de reducir la base de datos a un número computacionalmente manejable, manteniendo una estructura adecuada para el análisis que se iba a llevar a cabo, en cuanto a conexiones genealógicas, número de PDC por animal y tamaño de los niveles del grupo de comparación Rebaño-Fecha de control (RFC). Para mantener todos las PDC de los animales, manteniendo simultánea-

mente tamaños de RFC con un número de observaciones suficientes, se tuvieron que muestrear rebaños completos. Además, para mantener conexiones genealógicas entre los animales de la base de datos reducida que iba a ser analizada, los rebaños tuvieron que ser muestreados de una única región, escogiéndose además sólo aquellas lactaciones pertenecientes a un periodo de tiempo determinado. Tras esta primera selección se obtiene un archivo con 65 rebaños que es sometido a un proceso de depuración, en el que se exige una edad al parto entre 18 y 40 meses, un intervalo entre el parto y el primer control entre 5 y 67 días y un intervalo entre controles sucesivos entre 26 y 67 días. También se exige que todas las PDC estén comprendidas entre los 4 y 99 kg, y que se hayan registrado entre los 5 y 335 días de la lactación. Tras este proceso se obtuvo una muestra de 150.291 PDC correspondientes a 16.166 lactaciones, con 6.258 niveles del grupo RFC que mantenían 5 o más observaciones en cada uno de sus niveles. La duración media de la lactación fue de 291,83 ± 36,37 días. Sobre esta base de datos se muestrean aleatoriamente rebaños completos para obtener una muestra que será la base del análisis. Posteriormente se eliminan "outliers" y se consolida la muestra, exigiéndose que tenga un número mínimo de 6 PDC por animal, a la vez que se mantienen tamaños mínimos de 5 observaciones por nivel de RFC. Tras este proceso queda una muestra cuyas características se resumen en el cuadro 1.

Una de las técnicas de comparación de modelos que se van a utilizar es la validación cruzada (VC). Este método consiste en partir la muestra en dos subconjuntos. El primero es utilizado para estimar los parámetros del modelo, utilizando posteriormente estas estimas para predecir las PDC del segundo subconjunto. De esta forma podemos estudiar la capacidad predictiva del modelo. La partición se realiza de forma aleatoria bajo ciertas restricciones, debido a que para realizar las predicciones sobre los elementos del segundo subconjunto (PRE), es necesario que todos los niveles de lo efectos aquí contenidos lo estén también en el primer subconjunto (EST). Se realiza por tanto un muestreo aleatorio de 2 PDC por animal para formar el conjunto PRE, exigiéndose que todos los niveles de PRE estén a su vez en EST. Adicionalmente también se exige que los niveles de los RFC del conjunto EST tengan un tamaño mínimo de 3 observaciones. El número de PDC obtenidas tras el proceso de partición de la muestra para VC fue de 35.614 PDC para el conjunto EST con un 78,89% de las PDC contenidas en la muestra original, y de 9.536 PDC en el conjunto PRE co un 21,11% de los datos iniciales. En el

Cuadro 1
Resumen de datos de la muestra de datos analizada
Table 1
Summary of the data analysed

Rebaños	22
Lactaciones (animales en dato)	4.768
Numero PDC (≥ 6 PDC /animal)	45.150
Numero niveles RFC (≥ 5 PDC/nivel)	1.853
Animales en Genealogía	8.415

cuadro 2 se muestran las producciones medias y desviaciones estándar para las PDC y días en lactación (DEL) para el conjunto EST. Todos los parámetros de todos los modelos han sido estimados con la muestra de datos del conjunto EST, utilizándose únicamente los datos de los conjuntos PRE para la metodología de VC.

Se llevó a cabo la estimación de parámetros genéticos para PDC bajo distintos modelos de regresión aleatoria (MRA) y un modelo animal con medidas repetidas (MMR). El distinto comportamiento de los modelos fue evaluado en términos de la bondad de ajuste y de la capacidad de predicción. Las comparaciones también estuvieron basadas, en el caso de los MRA, en la habilidad de cada MRA para modelizar las estructuras de varianzas de los efectos aleatorios a lo largo de la lactación. Aunque es bien conocido el hecho de que son mejores aquellos modelos que permiten heterogeneidad de la varianza residual a lo largo de la lactación (OLORI y col., 1999a; OLORI y col., 1999b), hemos trabajado con varianza residual constante debido a que el programa utilizado en la estimación de los componentes de varianza no permite de momento la heterogeneidad en la varianza residual (MISZTAL, 2000). OLORI y col (1999a) afirman que mantener homogeneidad en la varianza residual a lo largo de la lactación, produciría subestimaciones de la varianza residual al principio y final de la lactación sin afectar de modo significativo a la estima del resto de componentes, por lo que pensamos que las conclusiones que se puedan extraer del presente trabajo son perfectamente válidas.

Modelos

Todos los modelos utilizados incluyen en su ecuación los mismos efectos fijos. Las

Cuadro 2
Media y desviaciónes estándar (paréntesis) para los días en lactación (DEL) y kg. de leche del conjunto EST
Table 2
Averages and standard deviations (in parenthesis) for days in milk (DEL) and milk yield of EST subset

Periodo	Intervalo (días)	DEL medio	Producción media (kg)	Número PDC
1	5-35	20,33 (8,83)	25,18 (6,21)	3.428
2	36-65	50,23 (8,68)	27,43 (6,49)	3.492
3	66-95	80,46 (8,73)	27,15 (6,66)	3.494
4	96-125	110,42 (8,61)	26,70 (6,69)	3.432
5	126-155	140,27 (8,55)	25,70 (6,73)	3.548
6	156-185	170,17 (8,66)	24,77 (6,72)	3.563
7	186-215	200,36 (8,61)	24,08 (6,76)	3.461
8	216-245	230,43 (8,64)	23,02 (6,59)	3.415
9	246-275	260,18 (8,55)	22,08 (6,52)	3.196
10	276-305	289,49 (8,57)	21,33 (6,50)	2.559
11	306-335	318,82 (8,64)	20,49 (6,40)	1.986

diferencias entre modelos se deben únicamente a si se utilizan o no regresiones aleatorias y en cuanto a las funciones que se han utilizado para definir esas regresiones.

Uno de los efectos fijos que vamos a tener en cuenta es el punto de la lactación en el que se ha producido el dato, mediante la inclusión de covariables que definen la forma general de la curva de lactación dentro del grupo de comparación edad-época de parto (EE). Para definir estas covariables, vamos a utilizar basándonos en análisis preliminares, la función lineal de ALI-SCHAEFFER de 5 covariables (ALI-SCHAEFFER, 1987), que define la producción de leche en el día *t* mediante la ecuación:

$$y(t) = b_0 + b_1\, t/305 + b_2 ln(t/305)^2 + \\ + b_3 ln(305/t) + b_4\, (ln(305/t))^2 \qquad [1]$$

donde:

y(t) producción en el día *t*.

b_i coeficientes de regresión estimados con el modelo. Dividir por la constante 305 no es estrictamente necesario pero reduce la magnitud de la covariable.

Teniendo esto en cuenta, los efectos fijos se escribirán para todos los modelos como:

$$EF_{ij} = RFC_i + EE_j + \sum_{m=1}^{4} \beta_{jm} X_{klm} \qquad [2]$$

donde:

EF_{ij} representa los efectos fijos (comunes a todos los modelos).

RFC_i es el grupo de comparación rebaño-fecha de control *i*.

EE_j es el grupo de comparación edad-época de parto *j*.

β_{jm} es el coeficiente de regresión fijo *m* anidado a EE_j.

X_{klm} es la covariable m de la función lineal de Ali-Schaeffer [1] definida para la PDC_l del animal *k*.

Modelo de repetibilidad

La ecuación del modelo es la siguiente:

$$y_{ijkl} = EF_{ij} + a_k + p_k + e_{ijkl} \qquad [3]$$

donde:

y_{ijkl} es la PDC_l del animal *k*, perteneciente al grupo RFC_i y cuya curva de lactación fija esta anidada al grupo EE_j.

EF_{ij} son los efectos fijos definidos en (2).

a_k es el aditivo del animal *k*.

p_k es el efecto ambiental permanente asociado a todas las *PDC* de la vaca *k*.

e_{ijkl} es el efecto ambiental transitorio de la PDC_l.

En notación matricial, **y** será el vector que contiene las PDC observadas en todos los animales, ***b***, ***a***, ***p*** y ***e*** son vectores que contienen los efectos ambientales (RFC_i, EE_j y los coeficiente de regresión fijos $\beta_{jm\ (m = 1,4)}$ asociados a la función de ALI-SCHAEFFER), aditivo, permanente y residual. ***X***, ***Z*** y ***W*** son matrices de incidencia, incluyéndose en *X* las covariables $x_{klm\ (m = 1,4)}$ de la función de ALI-SCHAEFFER [1].

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wp} + \mathbf{e} \qquad [4]$$

Se asume la siguiente estructura de (co)varianzas:

$$\text{var}\begin{bmatrix}\mathbf{a}\\ \mathbf{p}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{P} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{E}\end{bmatrix}$$

donde **G** es la matriz de (co)varianzas genéticas entre los animales, definida como $\mathbf{G} = \sigma^2_a\mathbf{A}$, siendo σ^2_a la varianza genética aditiva y **A** la matriz de parentesco. **P** es una matriz diagonal que contiene la varianza del efecto permanente, $\mathbf{P} = \sigma^2_p\mathbf{I}$ y **E** es una matriz diagonal que contiene la varianza residual σ^2_e.

Modelos de regresión aleatoria

Los MRA utilizados se pueden agrupar en modelos que ajustan coeficientes de regresión aleatorios sobre covariables definidas a partir de funciones lineales que describen la curva de lactación, y en modelos que ajustan coeficientes sobre covariables definidas a través de una combinación lineal de polinomios ortogonales (polinomios de Legendre). Estos últimos surgen del procedimiento de obtención de funciones de covarianza (FC) como funciones reales de dos variables, a través de bases definidas por un conjunto de polinomios ortogonales dentro de un espacio funcional. (KIRKPATRICK y HECKMAN 1989; KIRKPATRICK y BULMER, 1990; APOSTOL, 1973)

La ecuación general del modelo es:

$$y_{ijkl} = EF_{ij} + a_k + \sum_{m=1}^{r} \alpha_{km} Z_{1,klm} + p_k + \sum_{m=1}^{s} \omega_{km} Z_{2,klm} + e_{ijkl} \quad [5]$$

donde a_k y p_k se definen igual que en [3] y α_{km} coeficiente *m* de regresión aleatorio anidado al animal *k*. Obsérvese que *m* puede variar desde *1* hasta *r*, siendo r el número de términos dependientes del tiempo de la función ajustada a la parte aditiva. Considerando el termino independiente, el número de coeficientes de regresión estimados, u *orden de ajuste*, será $r + 1$.

$Z_{1,klm}$ covariable m definida por la función ajustada sobre el efecto aditivo

ω_{km} coeficiente de regresión aleatorio *m* anidado a p_k. El orden ajustado sobre el efecto permanente será $s + 1$.

$Z_{2,klm}$ covariable *m* definida por la función que se ajusta al efecto permanente.

Siguiendo la misma notación que en [4] para el MMR, un MRA se puede escribir en forma matricial como:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_x \mathbf{a} + \mathbf{W}_x \mathbf{p} + \mathbf{e} \quad [6]$$

con ***y***, ***b***, ***e*** y ***X*** definidos como en [4], siendo a el vector que contiene los $r+1$ coeficientes de regresión aleatorios ($a_k + \alpha_{km\ (m=1,r)}$). $\mathbf{Z}_x$ es la matriz de incidencia que relaciona las PDC con el animal e incluye las covariables $Z_{1,klm\ (m\ =\ 1,r)}$ correspondientes a la función lineal ajustada al efecto aditivo. El vector p incluye los $s + 1$ coeficientes de regresión aleatorios para el efecto permanente ($p_k + \varpi_{km\ (m\ =\ 1,s)}$) y la matriz $\boldsymbol{W}_x$ es la matriz de incidencia del efecto permanente que incluye también las covariables $Z_{2,klm\ (m\ =\ 1,s)}$ que se definen para modelizar este efecto a lo largo de la lactación. La estructura de (co)varianzas asumidas en los MRA es:

$$\text{var}\begin{bmatrix}\mathbf{a}\\ \mathbf{p}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{P} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{E}\end{bmatrix}$$

donde **G** se define como $\mathbf{G} = \Sigma_a \otimes \mathbf{A}$, siendo Σ_a una matriz que contiene las (co)varianzas genéticas entre los $r+1$ coeficientes de regresión aleatorios que se ajustan sobre la parte aditiva, y **A** la matriz de parentesco. P se define como $\mathbf{P} = \Sigma_p \otimes \mathbf{I}$, donde Σ_p es una matriz que contiene la (co)varianza ambiental permanente entre los $s+1$ coeficientes de regresión aleatorios ajustados sobre el efecto permanente. **E** es una matriz diagonal que contiene la varianza residual σ^2_e, asumida constante a lo largo de toda la lactación por limitaciones del programa empleado en la estimación de los componentes de varianza.

MRA basados en funciones lactacionales

Las funciones que determinan las covariables sobre las que se establece la regresión aleatoria son funciones lineales que definen una curva de lactación. Se han comparado dos modelos.

Modelo de Ali-Schaeffer (MRA-A)

La función lineal utilizada para definir las covariables en función de los DEL es la función lineal de 5 covariables de ALI-SCHAEFFER definida en [1]. Se ajusta esta función en las partes aditiva y permanente.

Modelo del Wilmink (MRA-W)

El modelo propuesto por WILMINK (1987) es una función lineal de tres covariables que describe la producción de leche $y(t)$ en función del tiempo t como:

$$y(t) = b_0 + b_1 t + b_2 exp(-0.05 t).$$

Se ajustó esta función tanto al efecto aditivo como al permanente.

MRA basados en polinomios de Legendre (MRA-L)

Se utiliza la serie de orden n de polinomios de Legendre normalizados (KIRKPATRICK y HECKMAN, 1989, KIRKPATRICK y BULMER, 1990; KIRKPATRICK y col 1994; MEYER y HILL, 1997; APOSTOL ,1973) donde el término de orden n de la serie viene dado por:

$$P_n(x) = \frac{1}{2^n}\sqrt{\frac{2n+1}{2}}\sum_{m=0}^{(n/2)}(-1)^m\binom{n}{m}\binom{2n-2m}{j}x^{n-2m} \qquad [7]$$

siendo x el DEL estandarizado dentro del intervalo (-1,1) que es donde está definida la serie ortogonal de polinomios de Legendre, calculada a partir de la expresión

$$x = \frac{2(t - t_{min})}{t_{max} - t_{min}} - 1,$$

siendo t días en lactación, t_{min} día mínimo de la lactación y t_{max} día máximo de la lactación. En la expresión [7] *(n/2)* indica tomar la parte entera de n/2, y los términos del sumatorio en los que se verifique que *n-2m=0*, son independientes de x.

La notación que vamos a seguir para describir un modelo al que se ajustan polinomios de Legendre es:

$$y = EF + L_a(n) + L_p(m) + e \qquad [8]$$

donde *EF* son los efecto fijos definidos en [2], y $L_a(n)$ y $L_p(m)$ representan los polinomios de Legendre de orden n y m ajustados sobre la parte aditiva y permanente, respectivamente. Para simplificar todavía más la expresión (8), en lo sucesivo nos referiremos a un MRA-L de la siguiente forma:

$$\begin{array}{l} La(2) + Lp(3) = A2 \\ La(2) + Lp(3) = A3 \\ \cdots\cdots \quad \cdots\cdots \quad \cdots\cdots \\ La(7) + Lp(6) = D7 \end{array}$$

donde la letra se refiere al orden del efecto permanente (A = 3; B = 4; C = 5; D = 6) y el número al orden de ajuste sobre el efecto aditivo. Por *orden* entendemos el número de coeficientes del polinomio.

Comparación de modelos

Para comparar los distintos modelos se han utilizado estadísticos que definen la bondad de ajuste y la capacidad para predecir datos faltantes. Además, en los MRA se estudio la forma en la que predecían las varianzas aditiva y permanente diarias a lo largo de la lactación, así como el cociente entre la varianza aditiva diaria (VAD) entre la suma de la varianza total.

Las estimas de los componentes de varianza, soluciones de los efectos y cálculo de las verosimilitudes para todos los modelos se realizaron con el programa REMLF90 (MISZTAL, 2000).

Bondad del ajuste

Para cuantificar la bondad del ajuste del modelo se usaron dos estadísticos: el porcentaje de sesgo al cuadrado para la capacidad de ajuste (PSB-F) y la correlación empírica para la capacidad de ajuste (RHO-F) entre el valor observado y el ajustado por el modelo. El PSB-F se define como:

$$\text{PSB-F} = \frac{\sum_{i=1}^{n} (y_{ijkl} - \hat{y}_{ijkl})^2}{\sum_{i=1}^{n} (y_{ijkl})^2} \times 100 \qquad [9]$$

siendo n el número total de observaciones, y_{ijkl} la PDC_l observada en el animal k dentro de los grupos RFCi y EE_j e $\hat{y}_{ijkl}$ la PDC calculada por la ecuación del modelo a partir de las estimas de los efectos asociados a esa observación. Un mayor PSB-F evidenciaría un peor ajuste del modelo a los datos.

Dentro de los MRA-L existen modelos anidados que hacen posible determinar por medio de pruebas chi-cuadrado sobre el ratio de verosimilitudes, el orden del polinomio hasta donde se produce un incremento significativo en la mejora del valor de la verosimilitud del modelo.

Capacidad de predicción

Un método para comparar modelos es la validación cruzada (VC). Una vez estimados los parámetros del modelo, podemos predecir datos faltantes, es decir, datos que no han intervenido en la estima de los parámetros del modelo, y en concreto vamos a predecir los datos contenidos en el conjunto PRE. Comparando el dato que predice la ecuación del modelo con el dato real contenido en PRE, podemos tener una medida comparativa de la capacidad de ajuste del modelo usando los mismos estadísticos que en el apartado anterior, donde ahora la correlación y el porcentaje de sesgo al cuadrado se definen en función del valor predicho y el dato faltante (RHO-P y PSB-P).

Estimas de las varianzas diarias

Con MRA en los que las covariables están definidas por funciones lineales, es posible estimar funciones de covarianza que permiten tener en cuenta la estructura de (co)varianzas entre las PDC a lo largo de la lactación. Si $\mathbf{z}_i$ es el vector de covariables para un tiempo t_i y $\mathbf{z}_j$ lo es para otro t_j, y $\mathbf{K}_a$ es la matriz de (co)varianzas entre los coeficientes de regresión aleatorios aditivos del modelo, entonces la (co)varianza aditiva a lo largo de la lactación viene dada por la siguiente función de dos variables:

$$g(t_i,t_j) = \mathbf{z}_i' \mathbf{K}_a \mathbf{z}_j \qquad [10]$$

En el caso en que $i = j$, la superficie que define [10], y en la que se encuentran las (co)varianzas de las PDC en función del tiempo, se reduce a una función de una variable que representa la evolución de la varianza a lo largo del ciclo de lactación. La expresión [10] definida para el efecto aditivo y para el permanente, junto con la varianza residual, nos va a permitir el cálculo de parámetros genéticos a lo largo de la lactación. Los cálculos fueron llevados a cabo entre los días 5 y 335 de la lactación.

Resultados y discusión

Comparación de modelos

Los estadísticos para *bondad de ajuste* obtenidos dentro de las series de MRA-L,

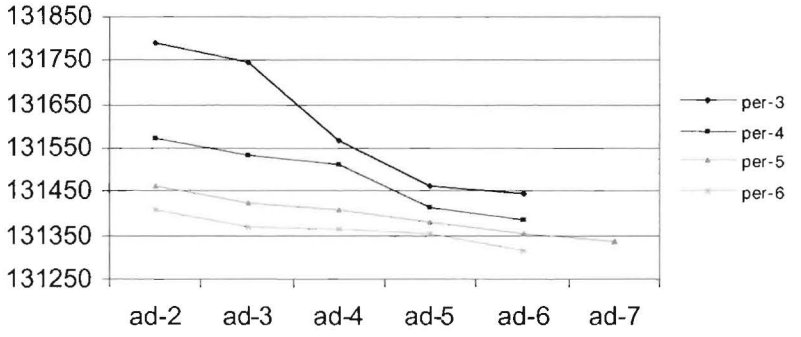


Figura 1. –2 log L*

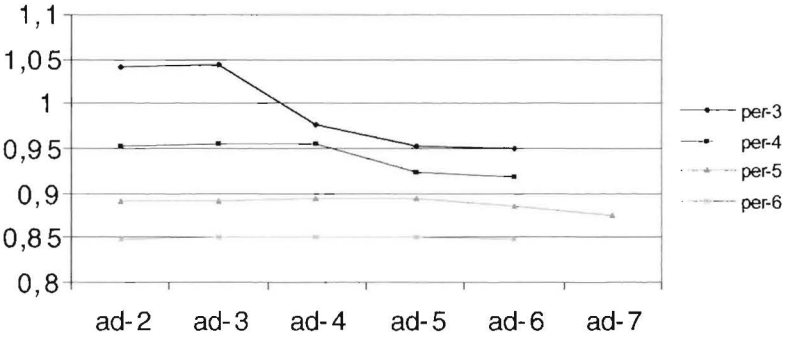


Figura 2. PSB para la bondad de ajuste- PSB-F
Figure 2. Goddness of fit using PSB- PSB-F

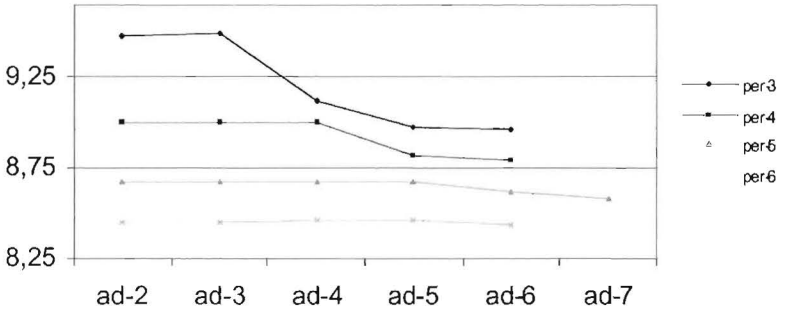


Figura 3. Varianza residual- var(e)*
*Figure 3. Residual variance- var(e)**

*En "ad-n" y "per-n", "n" indica el orden del polinomio que se ha ajustado a la parte aditiva y permanente respectivamente.
**In "ad-n" and "per-n", "n" indicates the order of the polynomial fitted to additive and permanent effect, respectively.*

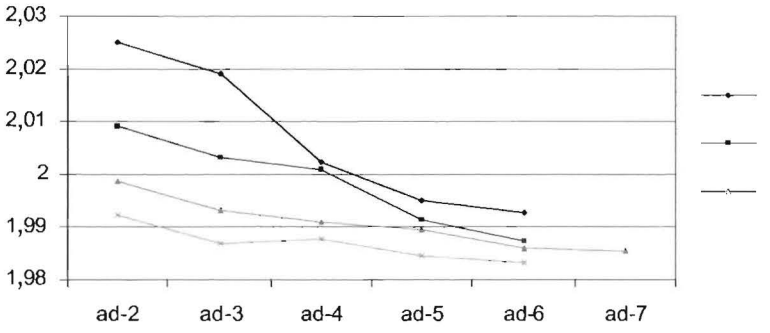


Figura 4. PSB para la capacidad predictiva PSB-P*
*Figure 4. PSB Predictive ability using- PSB-P**

*En "ad-n" y "per-n", "n" indica el orden del polinomio que se ha ajustado a la parte aditiva y permanente respectivamente.
**In "ad-n" and "per-n", "n" indicates the order of the polynomial fitted to additive and permanent effect, respectively.*

(logaritmo de verosimilitud, porcentaje de sesgo al cuadrado y varianza residual) se presentan en las figuras 1, 2 y 3 donde se muestra su tendencia ante el aumento de orden de ajuste sobre la parte aditiva dentro de distintos órdenes de ajuste constantes de la parte permanente. En el mismo sentido, la figura 4 muestra el porcentaje de sesgo al cuadrado para los MRA-L calculado en validación cruzada como medida de la *capacidad predictiva* de los modelos.

El valor de las verosimilitudes (figura 1) pone de manifiesto la mejora en la bondad de ajuste de los modelos paralela al incremento en el orden de los polinomios ajustados sobre los efectos aditivo y permanente. El valor de –2 log L disminuyó según se incremento el orden del polinomio sobre la parte aditiva manteniendo constante el orden de ajuste en la permanente, siendo la disminución en los valores de este estadístico más manifiesta en órdenes de ajuste sobre el efecto permanente inferiores (3 y 4) que en órdenes de ajuste superiores (5 y 6), donde las mejoras fueron más moderadas. Una tendencia similar en cuanto a la mejora del modelo, se encontró al analizar incrementos de orden sobre el efecto permanente dentro de un orden de ajuste constante sobre el efecto aditivo. Sin embargo, la disminución en el valor del –2 log L fue más pronunciada y se mantuvo para órdenes de efecto aditivo superiores (5 y 6). Sobre los ratios de verosimilitudes de los MRA-L anidados, se llevaron a cabo test chi cuadrado para determinar hasta que orden de ajuste se producía un descenso estadísticamente significativo en el valor de –2 log L (figura 1). En todos los casos resultó significativa la mejora que se produjo al incrementar el orden desde un grado 2 hasta grado 6, tanto en la parte aditiva como en la permanente. El paso de un modelo C6 a un C7 no resulto significativo. Los modelos de orden de ajuste 6 sobre el efecto permanente resultaron mejores que los de orden 5. El mejor modelo encontrado fue el D6. No se probó el modelo D7.

Otros estadísticos para medir la capacidad de ajuste, porcentaje de sesgo al cuadrado para la capacidad predictiva, (PSB-F) y varianza residual (var(e)), también informaron de que los mejores modelos eran los de órdenes superiores. En las figuras 2 y 3 se observa que la los estadísticos PSB-F y var(e) evolucionaron hacia valores cada vez más altos, al pasar de modelos con órdenes de ajuste inferiores sobre el efecto aditivo a modelos con órdenes superiores, mientras se mantuvo constante el orden de ajuste en el efecto permanente, hasta llegar a un determinado orden de ajuste en el que se produjo un descenso brusco en el valor de los estadísticos, que es suficiente para poner de manifiesto que los modelos de orden superior tienen mejores cualidades de ajuste. Este fenómeno no se puso de manifiesto si el orden del efecto permanente se aumenta sobre modelos con el efecto aditivo constante, en este caso, tanto PSB-F como var(e) disminuyeron a la vez que se incrementaba el orden del permanente. Dentro de los MRA-L, los valores de la correlación para la bondad de ajuste (RHO-F) mostraron el mismo comportamiento que el PSB-F y la var(e), oscilando entre un 92,76 y un 94,21%. En los estadísticos de medida sobre la capacidad de predicción, porcentaje de sesgo al cuadrado para la capacidad de predicción (PSB-P) (figura 4), y correlación (RHO-P), se encontró que mejo-

Cuadro 3
Estadísticos para los MRA-L más representativos de cara a la comparación, MRA lactacionales y el MMR (siglas de modelos explicadas en modelos de regresión aleatoria)
Table 3
Comparisons between MRA and MRR (model symbol explained in Random regression models section)

	Bondad del ajuste del modelo			Capacidad de predicción		
Modelo	-2 log L	PSB-F	RHO-F	var(e)	PSB-P	RHO-P
A2	131792	1,0424	0,9276	9,4842	2,0252	0,8553
A3	131744	1,044	0,9274	9,486	2,0191	0,8558
A4	131567	0,9774	0,9324	9,1277	2,0021	0,8571
A5	131461	0,9524	0,9343	8,981	1,9949	0,8576
B2	131571	0,9523	0,9343	9,001	2,0089	0,8566
B4	131514	0,9549	0,9341	9,0087	2,001	0,8572
B6	131389	0,9168	0,9369	8,7917	1,9874	0,8582
C2	131463	0,8908	0,9389	8,6715	1,9987	0,8574
C5	131381	0,8933	0,9387	8,6756	1,9897	0,858
C6	131356	0,8844	0,9394	8,6256	1,9857	0,8584
C7	131340	0,8751	0,9401	8,5761	1,9853	0,8584
D3	131370	0,8499	0,942	8,4537	1,9867	0,8583
D2	131408	0,8488	0,942	8,4511	1,9922	0,8579
D6	131316	0,8478	0,9421	8,436	1,9833	0,8585
MMR	134168	1,586	0,8859	12,493	2,2783	0,8354
MRA-W	131706	1,0658	0,9257	9,5713	2,03	0,8549
MRA-A		0,9755	0,9326	9,0957	2,0034	0,857

raban simultáneamente al incremento de orden de efectos aditivos y permanente. Los valores de RHO-P dentro de los MRA-L, oscilaron entre un 85,53 y un 85,85%. La variabilidad observada en los resultados de las medidas de capacidad de predicción fue menor que la observada en las de bondad de ajuste (cuadro 3).

Como era previsible, el MMR fue el modelo que presentó peores resultados, tanto desde el punto de vista de la bondad de ajuste como desde el de la capacidad predictiva. Los MRA que utilizaron funciones lactacionales para ajustarse a la parte aditiva y permanente, mostraron peores resultados que los MRA-L. A igualdad de número de parámetros ajustados, el modelo de Wilmink equivaldría a un MRA-L A3, presentando el modelo A3 mejores resultados. Por su parte, el MRA basado en la función lineal de Ali-Schaeffer, equivaldría en cuanto a número de parámetros a un MRA-L C5, siendo el modelo C5 mejor en cuanto a los resultados presentados. Un MRA de Ali-Schaeffer, se situaría en términos comparativos en una posición intermedia entre un A4 y un A5.

Estimas de parámetros genéticos

Se evaluaron las varianzas genética aditiva y permanente diarias (VAD y VPD) como funciones de los días en lactación (DEL). Los resultados para los modelos más representativos se muestran en las figuras 5 y 6.

La varianza genética aditiva estimada con el MMR fue de 5,80, mientras que la permanente y residual fue de 9,89 y 12,49, respectivamente. REKAYA y col. (1999) encuentran un valor similar para la varianza aditiva de 5,65, y valores algo más bajos para las varianzas permanente y residual de 6,60 y 6,68, respectivamente. Los MRA permiten oscilaciones de esta varianza a lo largo de la lactación. En la figura 5 se observa que dependiendo del MRA, el inicio de la curva es ligeramente distinto. Los modelos de Ali-

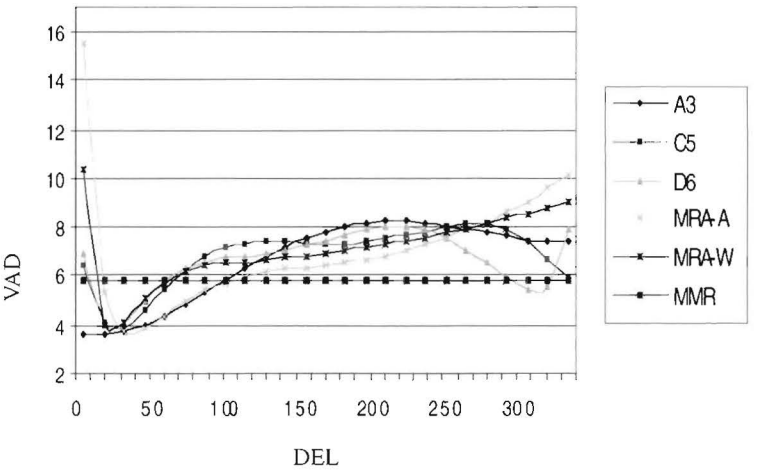


Figura 5. Varianza aditiva diaria- VAD *
*Figure 5. Daily additive variance- VAD **

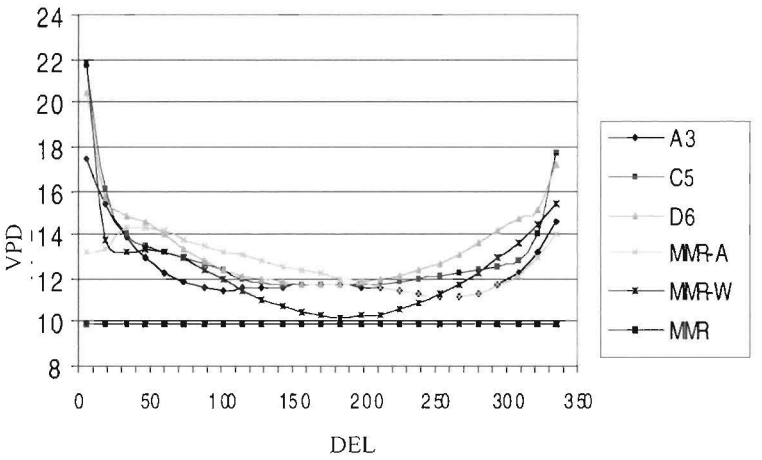


Figura 6. Varianza ambiental permanente diaria- VPD *
*Figure 6. Daily permanent environmental variance- VPD **

*Siglas de modelos explicadas en Modelos de regresión aleatoria.
** Model symbol explained in Random regression mode section.*

Schaeffer, Wilmink y los MRA-L de orden elevado, D6 y C5, toman valores altos al principio de la lactación decreciendo hasta el día 50. En el día 50 todos los modelos alcanzan el mínimo de VAD, y desde aquí todos muestran una línea ascendente hasta los 275-300 días. A partir de aquí cada modelo difiere en su comportamiento. En la figura 6 se ve como la VPD fue modelizada por todos los modelos de forma muy similar, presentando el MMR el valor más bajo.

La tendencia general de la "heredabilidad" (h^2) (figura 7) fue similar al de la VAD. Al principio de la lactación los modelos presentaron unos valores de h^2, relativamente altos, que fueron decreciendo hasta el día 50 con valores en torno a 0.15. Desde aquí aumentaron a valores en torno a 0.27, manteniéndose en la mayor parte de la lactación. A partir del día 300 los modelos manifestaron comportamientos distintos. Estas tendencias de la h^2 al principio y final de la lactación, se ponen de manifiesto en MRA que ajustan un elevado número de parámetros en al parte aditiva, por lo que parece ser debido al propio funcionamiento de los MRA. STRABEL y MISZTAL (1999) sugieren que este fenómeno se debe a la pérdida de precisión en la estima de los componentes de varianza que se produce los MRA cuando se incrementa el número de parámetros que han de ser estimados.

Este comportamiento también se puso de manifiesto en las "correlaciones genéticas diarias" (CGD). A partir de un orden cuarto de ajuste sobre la parte aditiva se obtuvieron correlaciones muy bajas e incluso negativas entre PDCs situados en la periferia de la lactación. Este resultado se muestra para el modelo D6 en la figura 8, habiéndose encontrado los mismos resultados para órdenes inferiores de efecto permanente desde órdenes de ajuste 4 para el efecto aditivo. Con órdenes 2 y 3 de ajuste sobre la parte aditiva se encontró que la correlación entre dos

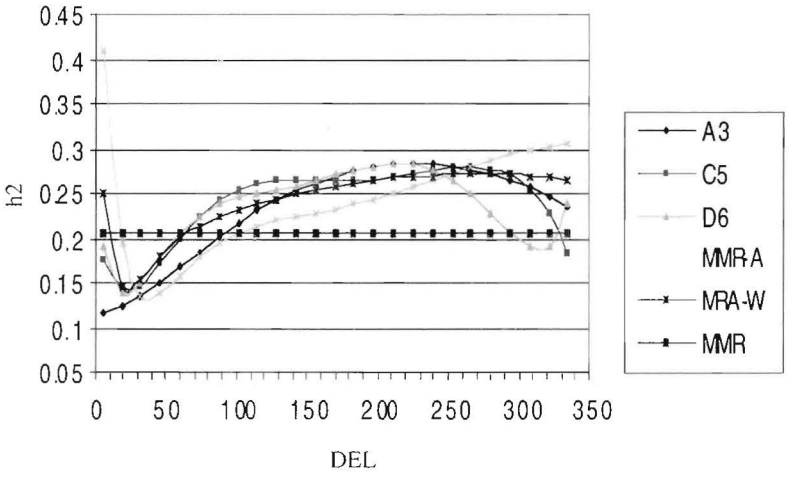


Figura 7. "Heredabilidad" diaria* (*Siglas de modelos explicadas en Modelos de regresión aleatoria)
Figure 7. Daily "heritability" (*Model symbol explained in Random regression models section)*

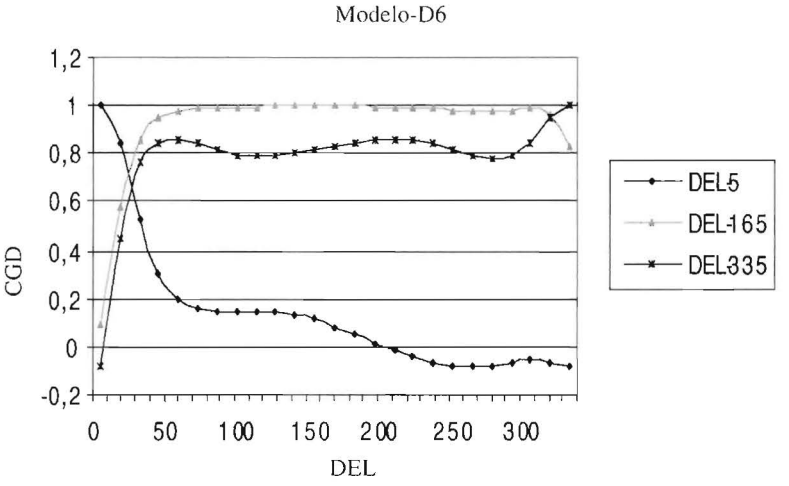


Figura 8. CGD para el modelo MRA-L D6
Figure 8. GDC for the MRA-L D6 model

PDCs medidos tiempos t_1 y t_2 disminuye a medida que t_2 aleja de t_1 (figura 9), encontrándose valores de CGD con un modelo D3 para PDCs entre los DEL 5 y 335 de 0,58. Resultados similares han sido presentados por STRABEL y MISZTAL (1999).

De igual forma los MRA-A y MRA-W mostraron resultados en las estimas de parámetros genéticos equivalentes a los MRA-L de elevado orden de ajuste (figuras 7, 10 y 11). En este caso, además del elevado número de parámetros del modelo (MRA-A), el

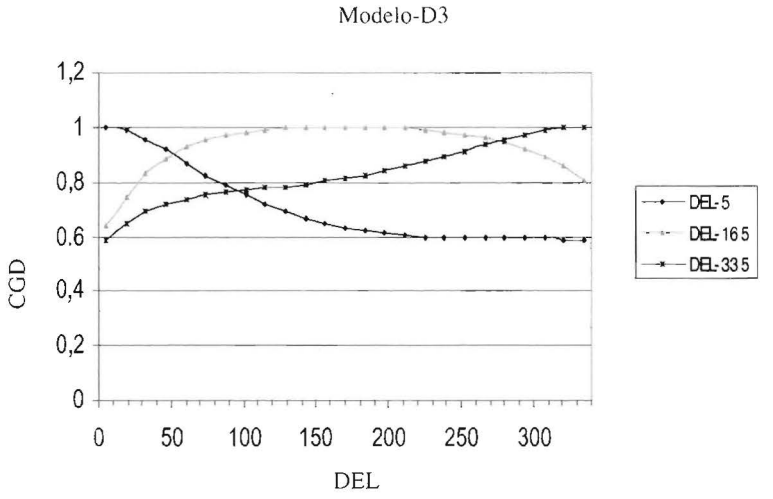


Figura 9. CGD para el modelo MRA-L D3
Figure 9. CGD for the MRA-L D3 model

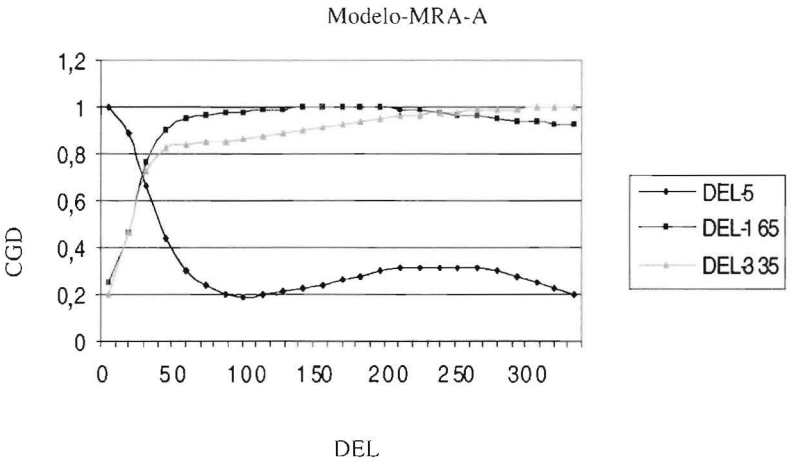


Figura 10. CGD para el modelo MRA-A
Figure 10. *CGD for the MRA-A model*

hecho de haber obtenido h^2 muy altas al principio y final de la lactación y haber obtenido CGD muy bajas para PDCs periféricos, también pudo ser debido a que los modelos lactacionales no son adecuados para modelizar al estructura de (co)varianzas, aunque sí lo sean para definir la curva de lactación.

Pequeñas imprecisiones en las estimas de los componentes de varianza del MRA, parecen traducirse en un comportamiento anómalo de los parámetros genéticos al principio y final de la lactación. En nuestro caso, las altas estimas encontradas para la VAD (figura 5) al principio de la lactación, tam-

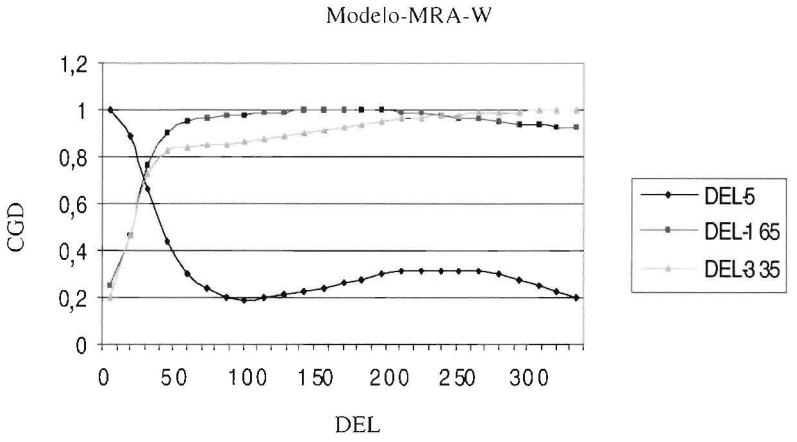


Figura 11. CGD para el modelo MRA-W
Figure 11. CGD for the MRA-W model

bién pudieron ser debidas a trabajar con modelos con homogeneidad en la varianza residual, sin embargo las estimas altas de la VAD sólo se manifestaron en MRA que ajustaron órdenes superiores a 3 en la parte aditiva. Este fenómeno no se observó si el incremento de orden se producía en la función que se ajustaba al efecto permanente. En este sentido, OLORI y col. (1999_a) afirman que mantener homogeneidad en la varianza residual a lo largo de la lactación en MRA, produciría sesgos en la propia varianza residual al principio y final de la lactación, pero no afectaría significativamente a las estimas del resto de las varianzas.

Los resultados obtenidos demuestran que aumentar el orden de los polinomios se traducen siempre en mejoras en la capacidad de ajuste, sin embargo esto lleva aparejado una mala modelización de los parámetros genéticos en la periferia de la lactación, lo que puede ser debido a una pérdida de precisión en la estima de los componentes de varianza cuando estos son incrementados en número, o al propio comportamiento del MRA que no se ajusta bien en los extremos de la lactación. Además, PLETCHER y GEYER (1999) afirman que los polinomios de alto grado son "oscilantes" por lo que no se ajustarían bien a la FC. Una solución a esto sería trabajar con polinomios de menor grado de ajuste. PLETCHER y GEYER señalan a su vez, que un polinomio de grado excesivamente reducido podría no ser suficiente para definir satisfactoriamente el proceso biológico que se este modelizando. Por tanto, parece lógico que hay que definir un polinomio con un orden de ajuste óptimo, para lo que habría que tener en cuenta, además de factores como la capacidad de ajuste y la capacidad para predecir datos faltantes, otros relacionados con la precisión de los componentes de varianza y la posterior estima de parámetros genéticos, teniendo especial atención con lo que ocurre en los extremos de la lactación, con la capacidad del polinomio para definir el proceso biológico, y finalmente con el coste computacional.

Conclusiones

Los resultados de la comparación de modelos obtenidos, tanto en lo referente a bondad de ajuste como capacidad de predicción, discriminaron muy en favor de los MRA frente al MMR.

Dentro de los MRA, los MRA-L mostraron mejores resultados que los MRA basados en funciones lactacionales (MRA-A y MRA-W), a igualdad de número de parámetros ajustados.

En los MRA-L, se encontró que los modelos mejoraron sus propiedades en cuanto a bondad de ajuste y capacidad de predicción de forma paralela al aumento en el número de parámetros, precisando que, incrementos en el orden del polinomio que se ajustó sobre la parte aditiva, produjeron mejoras más acusadas en modelos donde el orden ajustado sobre la parte permanente no fue muy elevado (3-4). Sin embargo, los mismos incrementos en el orden de ajuste sobre la parte aditiva en modelos con un orden elevado en la permanente (5-6) produjeron mejoras poco relevantes. Por su parte, incrementar el orden del efecto permanente produjo mejoras en el modelo independientemente del orden que se ajustó sobre la parte aditiva. Por tanto, desde el punto de vista del ahorro computacional, parece razonable tratar de mejorar las características del modelo incrementando el orden de ajuste sobre la parte permanente manteniendo órdenes de ajustes no muy elevados en la parte aditiva.

Por otro lado, el aumento en el número de parámetros del modelo condujo a resultados muy elevados de h^2 al principio y final de lactación, a la vez que resultados de CGD muy bajos entre PDCs situadas en la periferia de la lactación. No obstante, este efecto sólo se puso de manifiesto cuando el incremento de orden de ajuste se produjo sobre la parte aditiva del modelo.

Sumado a esto último el hecho de que, las mejoras que se produjeron en el modelo, como resultado de incrementar el orden de ajuste en la parte aditiva en modelos de elevado orden de ajuste en la parte permanente no fueron muy superiores, nos podemos permitir sugerir que el modelo MRA-L D3 sería el más adecuado para su uso, por ser un modelo que presenta buenas características en cuanto a bondad de ajuste y capacidad de predicción, sin presentar el inconveniente de la imprecisión en la estima de los componentes de varianza.

Dado que los resultados obtenidos se basan en modelos que consideraron homogeneidad en la varianza residual, y es conocido el hecho de que modelos con heterogeneidad de la varianza residual presentan mejores resultados, se deberían hacer más investigaciones al respecto. Así mismo, se deberían comparar los resultados de este análisis, con otros que se obtengan a partir de otras muestras de datos en otras regiones. Por último, se debe plantear un estudio sobre el posible cambio en el mérito genético de los animales que se podría producir con modelos basados en el análisis de las PDC.

Bibliografía

ALI T.T., SCHAEFFER L.R., 1987. Accounting for covariances among Test day milk yield in dairy cows.

APOSTOL T.M., 1973. Calculus. Vol- 1. pag-700.

JAMROZIK J., KISTEMAKER J., SCHAEFFER L.R., DEKKERS J. C.M., 1997. Comparison of possible covariates for use in a random regression model for analyses of test day yields. J. Dairy Sci. 80: 2550.

KIRKPATRICK M., HECKMAN N., 1989. A quantitative genetic model for growth, shape, reaction norms,

and other infinite-dimensional characters. J. of Math. Biol. 27: 429.

KIRKPATRICK M., BULMER M., 1990. Anaysis of the inheritance, selection and evolution of growth trajectories. Genetics. 124:979.

KIRKPATRICK M., HILL W.G. y THOMPSON R., 1994. Estimationg covariance structure of traits during growth and ageing, illustration with lactations in dairy cattle. Genet. Res. 64:57.

MEYER K., HILL W.G., 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or "repeated" records by restricted maximum likelihood. Livest. Prod. Sci. 47:185.

MISZTAL I., 2000. REMLF90 Manual. ftp://num.ads.uga.edu

OLORI V.E., HILL W.G., BROTHERSTONE S., 1999a. The structure of the residual error variance of test day milk yield in random regression models. Workshop on Computational Breeding. Finland. 18-20 March.

OLORI V.E., HILL W.G., MCGUIRK, BROTHERSTONE S., 1999b. Estimating variance components for test day milk records by restricted maximum likelihood with a random regression model. Livest. Prod. Sci. 61: 53.

PLETCHER S.D., GEYER C.J., 1999. The Genetic analysis of age-dependente traits: Modeling the character process. Genetics: 151:825.

PTAK E., SCHAEFFER L., 1993. Use of test day yields for genetic evaluation of dairy sires and cows. Livest. Prod. Sci. 34: 23.

REENTS R., JAMROZIK J., DEKKERS J.C.M., SCHAEFFER L.R., 1995a. Estimation of genetic parameters for test day records of somatic cell score. J. Dairy Sci. 78:2847.

REENTS R., DEKKERS J.C.M., SCHAEFFER L.R., 1995b. Genetic evaluation for somatic cell score with a test day model for multiple lactations. J. Dairy Sci. 78:2858.

REKAYA R., CARABAÑO M.J., TORO M.A., 1999. Use of test day yields for the genetic evaluation of production traits in Holstein-Friesian cattle. Livest. Prod. Sci. 57:203.

STRABEL, T., MISZTAL I., 1999. Genetic parameters for first and second lactation milk yields of Polish black and white cattle with random regression test-day model. J Diary Sci. 82:2805.

TIJANI A., WIGGANS G.R. , VAN TASSEL C.P., PHILPOT J. C., GENGLER N., 1999. Use of (co)variance functions to describe (co)variances for test day yield. J. Dairy Sci. 82.

VAN DER WERF, J.H.J., GODDARD M.E., MEYER, K., 1998. The use of covariance funtion and random regressions for genetic evaluation of milk production based on test day records. J Diary Sci. 81: 3300.

WILMINK J.B.M., 1987. Adjusment of test day milk, fat and protein yield for age, season ans stage of lactation. Livest. Prod. Sci. 16:335.

ITEA (2000), Vol. 96A N.° 3, 337-348

ANÁLISIS DE LA RESPUESTA CORRELACIONADA A LA SELECCIÓN POR VELOCIDAD DE CRECIMIENTO EN LAS CURVAS DE CRECIMIENTO DE CONEJOS

M. Piles*, D. Gianola, L. Varona***, A. Blasco***

* Dpto. de Ciencia Animal-Universidad Politécnica de Valencia, 46071 Valencia, España

** Department of Animal Science, University of Wisconsin, Madison, Wisconsin 53706, USA

*** Centre UdL-IRTA, 25198 Lérida, España

Introducción

La mayoría de los esquemas de selección para producción de carne en conejo incluyen una línea seleccionada por velocidad de crecimiento. La selección por velocidad de crecimiento puede alterar la pauta de crecimiento de los animales cambiando el peso al destete, la edad a la cual es alcanzado el peso de sacrificio o el peso en el estado adulto. La trayectoria de crecimiento puede ser descrita mediante alguna función (GOMPERTZ, LOGÍSTICA, RICHARDS, etc.) cuyos parámetros poseen interpretación biológica y pueden ser usados para obtener algunas medidas de interés económico o para detectar posibles cambios en la pauta de crecimiento de los animales como consecuencia de la selección por otro carácter correlacionado. La función de Gompertz ha sido frecuentemente utilizada en conejo (GÓMEZ y BLASCO, 1992; SEELAND *et al.*, 1996; FIORELLO y GERMAN, 1997) por presentar normalmente un buen ajuste a los datos y por estar definida por sólo 3 parametros (a,b,k). El parámetro a define el peso en el estado adulto, el parámetro b está relacionado con el peso al inicio del crecimiento y el parámetro k es un índice de decaimiento de la velocidad de crecimiento relativa o tasa de aproximación al estado adulto. No existe información previa acerca de la respuesta correlacionada a la selección en los parámetros que definen la curva de crecimiento en conejo. Los únicos trabajos que tratan de esta cuestión se basan en la comparación entre líneas por lo que las diferencias encontradas podrían deberse al diferente origen genético de las mismas.

Cuando no se dispone de una población control es necesario incluir en el análisis toda la información sobre la que se basaron las decisiones de selección para obtener inferencias correctas acerca de la respuesta correlacionada (GIANOLA y FERNANDO, 1986; GIANOLA *et al.*, 1999). El procedimiento propuesto por VARONA *et al.* (1997, 1998) y REKAYA (1997) basado en la utilización de un modelo jerárquico no lineal puede ser empleado cuando se dispone de datos longitudinales (registros del peso vivo a lo largo de la vida del animal) para la mayor parte de los animales a lo largo del proceso de selección utilizando la técnica bayesiana del aumento de datos para generar los residuos correspondientes a aquellos animales sin este tipo de datos. Toda la información del conjunto de datos es usada

a través del conocimiento de las relaciones genéticas entre individuos para estimar la media genética de una determinada generación. Sin embargo en la mayoría de los experimentos relacionados con el estudio de la pauta de crecimiento de los animales se dispone únicamente de datos longitudinales de un número reducido de animales en un momento determinado del proceso de selección debido al elevado coste que supone su obtención. En ese caso la aplicación del procedimiento anterior resulta impracticable pues el tiempo requerido para la convergencia sería excesivamente elevado. El objetivo de este trabajo es estimar la respuesta correlacionada a la selección por velocidad de crecimiento en la pauta de crecimiento de los animales en esta situación mediante la aplicación del procedimiento de análisis estadístico descrito en un trabajo previo (GIANOLA *et al.*, 1999) que permite superar esta dificultad de una manera sencilla.

Material y métodos

Animales

Los animales proceden de una línea sintética de conejos seleccionada mediante selección individual por incremento en la ganancia media diaria (adg) de peso en el periodo comprendido entre las 4 y las 9 semanas de vida (ESTANY *et al.*, 1992). Esta línea fue creada mediante el cruzamiento de animales procedentes de una línea de origen californiano con los de una línea sintética creada mediante apareamientos entre animales procedentes de diversas líneas comerciales de machos terminales. No se practicó selección durante las dos primeras generaciones con objeto de reducir el desequilibrio gamético. Las generaciones eran discretas. Para reducir el incremento de la consanguinidad no se realizaron apareamientos entre animales con abuelos comunes y los machos eran seleccionados dentro de familia paterna.

Los animales permanecieron alojados en la granja de la Universidad Politécnica de Valencia que dispone de mecanismos de control de luz, temperatura y ventitación. Tras el destete los gazapos pasaban a ocupar jaulas de 8 individuos y eran alimentados a voluntad con un pienso comercial (16,0% proteína cruda, 15,5% fibra, 3,4% grasa) hasta alcanzar las 9 semanas de vida. Posteriormente eran alojados en jaulas individuales y alimentados con el mismo pienso cuyo consumo era restringido a 140 g/d. A las 20 semanas de edad los animales eran llevados a la nave de maternidad donde se alojaban en jaulas individuales y eran alimentados con un pienso comercial con 17,5% proteína cruda, 14,5% de fibra y 3,4% de grasa.

Una muestra aleatoria de 76 animales de la generación 11 de selección fueron individualmente pesados dos veces por semana durante las 4 primeras semanas de vida y una vez por semana desde este momento hasta las 40 semanas en los machos y las 20 semanas en las hembras. Por otra parte embriones pertenecientes a las generaciones 3 y 4 de selección fueron descongelados y transferidos a hembras receptoras. Los animales nacidos de estas transferencias (6 animales de la generación 3 y 4 de la generación 4) eran contemporáneos de los animales de la generación 10 de selección. Para evitar el posible efecto de la congelación en el crecimiento de los animales se realizaron apareamientos entre los animales de la generación 3 con los de la generación 4 siendo los animales resultantes de estos apareamientos contemporáneos de los de la generación 11 de selección, los cuales también fueron pesados del mismo modo que aquellos. Estos animales (61) constituían el

grupo control de un trabajo previo en el que la respuesta correlacionada a la selección en los parámetros de la curva de crecimiento era estimada mediante la comparación de este grupo con el grupo de animales de la generación 11 de selección. Los datos correspondientes a este grupo control fueron incluidos en este análisis con objeto de mejorar la precisión de las estimas.

Modelo estadístico y análisis

Estructura de los datos

Se disponía de un conjunto

$$B = \left\{B_L, B_{\bar{L}}\right\}$$

de $N = 10{,}151$ individuos pertenecientes a las distintas generaciones a lo largo del proceso de selección de los que se conoce la velocidad de crecimiento durante el periodo de cebo. De ellos el subconjunto B_L de $M = 137$ (76+61) animales poseían datos longitudinales (L) siendo B_L el subconjunto complementario consistente de $N - = 10{,}014$ individuos. El vector de datos correspondiente a los animales pertenecientes a B_L será denotado por el vector

$$y_i = \begin{bmatrix} y_i^D \\ y_i^C \end{bmatrix},$$

de orden $(n_i + 1)x1$; $i \in B_L$, donde

$$y_i^C = \left\{y_{ij}\right\};\quad i = 1,2,...,137;\quad j = 1,2,...n_i$$

es el vector de n_i pesos registrados a diferentes edades en el individuo i^{th} ($n_i \leq 44$) e y_i^D es su velocidad de crecimiento. Para los anmales pertenecientes al subconjunto B_L, la única información disponible era el criterio de selección y_i^D.

Se dispuso de un total de 3,527 registros de pesos individuales y había 10,159 animales en la genealogía.

Los datos y los parámetros del modelo relacionados con el criterio de selección serán denotados con el superíndice "*D*" mientras que los relacionados con los parámetros que definen la curva de crecimiento de los animales se denotarán por el superíndice "*C*".

Métodos

Se realizó un análisis bayesiano basado de un modelo jerárquico no lineal. En el primer estrato de la jerarquía se define el modelo que describe la variación del peso a lo largo del tiempo. Así la observación del individuo *i* a la edad *j* viene dada por la expresión:

$$y_{ij}^C = f_{ij}\left(\theta_i, t_{ij}\right) + \varepsilon_{ij} =$$

$$= a_i \cdot \exp\left[-b_i \cdot \exp\left(-k_i \cdot t_{ij}\right)\right] + \varepsilon_{ij}$$

siendo $\theta_i' = \left\{a_i, b_i, k_i\right\}$ el vector de parámetros de la función de Gompertz específicos de ese animal. El vector de pesos registrados a lo largo de la vida del animal es expresado entonces como:

$$y_i^C = f_i\left(\theta_i, t_i\right) + \varepsilon_i;\quad i \in B_L$$

siendo $f_i\left(\theta_i, t_i\right)$ el vector de pesos esperados de acuerdo con la función de Gompertz, t_i el vector de orden $n_i x\ 1$ de edades a las cuales fueron registrados los pesos y, ε_i el vector de residuos de orden $n_i x\ 1$.

Se asume que los datos se distribuyen normalmente y que los residuos son independientes entre individuos y dentro de un mismo individuo de manera que la matriz $R_i(\gamma)$ de orden $n_i x n_i$ de (co)varianzas residuales entre los pesos del individuo *i* es una matriz diagonal cuyos elementos son función de un vector de parámetros γ:

$$y_i^C | \theta_i, \gamma \sim N\left[f_i\left(\theta_i, t_i\right), R_i(\gamma)\right]$$

El vector $\gamma = \{a_\varepsilon, b_\varepsilon, k_\varepsilon\}$ contiene los parámetros de una función de Gompertz ya que se observó que debido a un efecto de escala, la evolución con el tiempo de la desviación típica residual podía ser bien descrita por una función de crecimiento. Se asume que todos los animales tienen la misma desviación típica residual a la misma edad y así a una edad *j*:

$$\sigma_j = f(\gamma, t_j) = a_\varepsilon \cdot \exp\left[-b_\varepsilon \cdot \exp\left(-k_\varepsilon \cdot t_j\right)\right] \forall$$
$$j = 1,2,...,44$$

El vector completo de datos longitudinales se expresa como:

$$y^C = f(\theta, t) + \varepsilon$$

donde θ es el vector de parámetros de todos los animales, de orden *(M x 3) x 1* y, ε es el vector de residuos de orden

$$\left(\sum_{i=1}^{M} n_i\right) x 1$$

La asunción de independencia condicional conduce a que la distribución de los datos dados los parámetros pueda ser expresada como:

$$p\left(y^C|\theta, \gamma\right) = \prod_{i \in B_L} \prod_{j=1}^{n_i} p\left(y_{ij}^C|\theta_i, \gamma\right) =$$

$$= \prod_{i=1}^{M} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\sigma_j} exp\left[-\frac{\left[y_{ij}^C - a_i exp\left\{-b_i exp(-k_i t_{ij})\right\}\right]^2}{2\sigma_j^2}\right]$$

En el segundo estrato del modelo se describen las causas de variación entre individuos en la velocidad de crecimiento (y^D) los parámetros que definen la trayectoria individual de crecimiento (θ). Se utilizó el siguiente modelo:

$$y_i^D = x'^D_i \theta^D + u_i^D + c_i^D + e_i^D; \text{ for all } i \in B_L$$

$$\begin{bmatrix} y_i^D \\ \theta_i \end{bmatrix} = \begin{bmatrix} x'^D_i & 0 \\ 0 & X_i^C \end{bmatrix} \begin{bmatrix} \beta^D \\ \beta^C \end{bmatrix} + \begin{bmatrix} u_i^D \\ u_i^C \end{bmatrix} + \begin{bmatrix} c_i^D \\ 0_i^C \end{bmatrix} + \begin{bmatrix} e_i^D \\ e_i^C \end{bmatrix};$$

for all $i \in B_L$

en el que:

- β^D es el vector que incluye los efectos de año-estacion (26 niveles), orden de parto (3 niveles: 1.º, 2.º, 3.º o posterior) y tamaño de camada (8 niveles: <6,6,7,8,9,10,11,>11) en la velocidad de crecimiento, y_i^D.
- β^C es el vector que incluye el efecto del sexo sobre los parámetros de la curva de crecimiento θ_i.
- u_i^D y u_i^C son los efectos genéticos aditivos para la velocidad de crecimiento y los parámetros de la curva de crecimiento del animal *i* respectivamente.
- c es el efecto ambiental de camada de origen sobre la velocidad de crecimiento del individuo *i*.
- x'^D_i y X_i^C son el vector y la matriz de incidencia que relacionan los datos con los efectos incluidos en los vectores β^D y β^C respectivamente.
- e_i^D and e_i^C son los residuos.

Se asumió que los datos se distribuían normalmente y que los residuos eran independientes entre individuos pero no dentro de un mismo individuo:

$$y_i^D|\beta^D, u_i^D, c_i^D, \sigma_D^2 \sim N\left(x'^D_i \beta^D + u_i^D + c_i^D, \sigma_D^2\right);$$

for all $i \in B_L$

$$\begin{bmatrix} y_i^D \\ \theta_i \end{bmatrix} \Big| \beta^D, \beta^C, u_i^D, c_i^D, u_i^C, \Sigma_0 \sim N\left(\begin{bmatrix} x'^D_i & 0 \\ 0 & X_i^C \end{bmatrix} \begin{bmatrix} \beta^D \\ \beta^C \end{bmatrix} + \begin{bmatrix} u_i^D \\ u_i^C \end{bmatrix} + \begin{bmatrix} c_i^D \\ 0 \end{bmatrix}, \Sigma\right); \text{ for all } i \in B_L$$

Donde Σ_0 es la matriz de (co)varianzas entre los residuos del segundo estrato del modelo jerárquico, de orden (4 x 4). Para los individuos con datos longitudinales y el correspondiente al criterio de selección se tiene que,

$$\theta_i|y_i^D,\beta^D,\beta^C,u_i^D,c_i^D,u_i^C,\Sigma \sim N\left(X_i^C\beta^C + u_i^C + \right.$$

$$\left. + b_\Sigma\left(y_i^D - x_i^{D'}\beta^D - u_i^D - c_i^D\right),\Sigma_{C.D}\right)$$

$$\text{donde: } b_\Sigma = \frac{1}{\sigma_D^2}\begin{bmatrix}\sigma_{Da}\\ \sigma_{Db}\\ \sigma_{Dk}\end{bmatrix}$$

es un vector de coeficientes de regresión y,

$$\Sigma_{C.D} = \left\{\sigma_l^2 - \frac{\sigma_{lD}^2}{\sigma_D^2};\ l=a,b,k \text{ and } \sigma_{lm} - \frac{\sigma_{Dl}\sigma_{Dm}}{\sigma_D^2};\right.$$

$$\left. l\neq m;\ l,m=a,b,k\right\}$$

operando se puede ver fácilmente que

$$\Sigma = \begin{bmatrix}\sigma_D^2 & b_\Sigma'\sigma_D^2\\ b_\Sigma\sigma_D^2 & \Sigma_{C.D} + b_\Sigma b_\Sigma'\sigma_D^2\end{bmatrix}$$

lo cual permite una parametrización alternativa del problema en función de b_Σ, $\Sigma_{C.D}$ y σ_D^2 en lugar de Σ con lo que ya no resulta necesario aumentar los residuos de aquellos animales sin datos longitudinales.

La densidad de la distribución a priori conjunta de todas las incognitas del sistema ($\beta^C,\beta^D,u^C,u^D,c^D,\Sigma,\gamma$) se asumió que era:

$$p\left(\beta^C,\beta^D,u^C,u^D,c^D,\Sigma,\gamma|G\right) =$$

$$p\left(\beta^C,\beta^D,u^C,u^D,c^D,\Sigma_{C.D},b_\Sigma,\sigma_D^2,\gamma|G\right) =$$

$$= p\left(\beta^C\right)\ p\left(\beta^D\right)\ p\left(u^C,u^D|G\right)\ p\left(c^D|\sigma_c^2\right)$$

$$p\left(\Sigma_{C.D}\right)\ p\left(b_\Sigma\right)\ p\left(\sigma_D^2\right)\ p\left(\sigma_c^2\right)\ p\left(\gamma\right)$$

donde G es la matriz de (co)varianzas genéticas de orden (4 x 4). Las distribuciones a priori fueron distribuciones uniformes acotadas para β^C, β^D, $\Sigma_{C.D}$, b_Σ, σ_D^2, σ_c^2, γ y G, y distribuciones normales para los efectos génicos aditivos y los efectos ambientales $\begin{bmatrix}u_C\\ u_D\end{bmatrix} \sim N\left(\begin{bmatrix}0\\ 0\end{bmatrix}, G\otimes A\right)$ donde A es la matriz de relaciones genéticas aditivas

$$\left[c^D\right] \sim N\left(0,\ I\sigma_c^2\right)$$

Para los individuos que poseen datos longitudinales y el dato del criterio de selección la densidad de la distribución conjunta de y_i^C,θ_i e y_i^D se puede escribir como:

$$p\left(y_i^C,\theta_i,y_i^D|\ \beta^C,\beta^D,u_i^C,u_i^D,c_i^D,\Sigma,\gamma\right) =$$

$$= p\left(y_i^D|\ \beta^D,u_i^D,c_i^D,\sigma_D^2\right)\ p\left(y_i^C|\ \theta_i,\gamma\right)$$

$$p\left(\theta_i|\ \beta^C,\beta^D,u_i^C,u_i^D,\Sigma,y_i^D\right)$$

ya que y_i^C y y_i^D son condicionalmente independientes porque los errores de ajuste de la trayectoria longitudinal son asumidos independientes entre individuos e independientes de los residuos del segundo estrato, dados los parámetros de la curva θ_i. Esto permite expresar la distribución posterior conjunta como:

$$p(\lambda\,|\,y^D,y^C) \propto p(y^D\,|\,\beta^D,u^D,c^D,\sigma_D^2)\ p(y^C$$

$$|\,\theta,\gamma)\ p(\theta\,|\,\beta^C,\beta^D,u^C,u^D,?_{C.D},b_\Sigma,y^D)$$

$$p(\beta^C,\beta^D,u^C,c^D,u^D,\Sigma_{C.D},b_\Sigma,\sigma_D^2,\sigma_c^2,\gamma,G)$$

$$p(\lambda\,|\,y^D,y^C)\propto$$

$$\propto\prod_{i=1}^{N}\left(\sigma_D^2\right)^{-\frac{1}{2}}\exp\left[-\frac{1}{2\sigma_D^2}\left(y_i^D - x_i^{D'}\beta^D - u_i^D - c_i^D\right)^2\right]\ x$$

$$\prod_{i=1}^{M}|R_i(\gamma)|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[y_i^C - f_i(\theta_i,t_i)\right]\right.$$

$$\left.\left[R_i(\gamma)\right]^{-1}\left[y_i^C - f_i(\theta_i,t_i)\right]\right\}\ x$$

$$\prod_{i=1}^{M}|\Sigma_{C.D}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[\theta_i - \left(X_i^C\beta^C + u_i^C + b_\Sigma\right.\right.\right.$$

$$\left.\left.\left.\left(y_i^D - x_i^{D'}\beta^D - u_i^D - c_i^D\right)\right)\right](\Sigma_{C.D})^{-1}\right.$$

$$\left.\left[\theta_i - \left(X_i^C\beta^C + u_i^C + b_\Sigma\left(y_i^D - x_i^{D'}\beta^D - u_i^D - c_i^D\right)\right)\right]\right\}$$

$$|G|^{-\frac{q}{2}}|A|^{-1}\exp\left(-\frac{1}{2}u'(G\otimes A)^{-1}u\right)$$

de la que fueron derivadas las distribuciones condicionales posteriores completas necesarias para la implementación del algoritmo

de muestreo de Gibbs el cual fue utilizado para aproximar las distribuciones marginales posteriores de las variables de interés. Las distribuciones condicionales posteriores completas tenían todas forma conocida excepto las correspondientes a los parámetros b y k de la función de Gompertz y los parámetros contenidos en el vector γ por lo que se utilizó el algoritmo de muestreo de Metropolis-Hastings utilizando como densidad generadora de candidatos distribuciones uniformes acotadas centradas en el valor actual del parámetro.

La respuesta directa y correlacionada a la selección fue estimada como la diferencia entre la media de los valores genéticos aditivos de los animales de la generación 11 de selección y la de los individuos de las generaciones 3 y 4 de selección para facilitar la comparación de los resultados con los obtenidos mediante la la comparación entre grupos en un análisis anterior.

Se obtuvieron 4 cadenas de longitud 555000, 790000, 747500 y 600000 muestras respectivamente, cada una de ellas partiendo de valores distintos para cada uno de los parámetros. Se aplicó el procedimiento de Ratery y Lewis (1992) para determinar la longitud del periodo de calentamiento. Este procedimiento se llevó a cabo utilizando las 22200, 31600, 29900 y 24000 muestras retenidas de cada cadena (1 de 25). Se aplicó el test de Gelman y Rubin basado en la comparación de la varianza entre cadenas y dentro de cada cadena para el diagnóstico de convergencia. Esta comparación es usada para estimar el factor por el cual el parámetro de escala de la distribución marginal posterior de cada variable podría ser reducido si la cadena tuviese una longitud infinita.

Resultados y discusión

En las figuras 1 a 4 se muestran las trazas superpuestas de cada una de las cadenas para la respuesta correlacionada a la selección en cada uno de los parámetros de la curva de Gompertz y para la respuesta directa. En ellos se puede apreciar la elevada correlación entre las muestras de cada uno de las cadenas y como consecuencia de esto el lento movimiento en el espacio paramétrico. Las trazas de las cadenas no se solapan completamente pero tampoco parecen permanecer en la región del espacio determinada por los

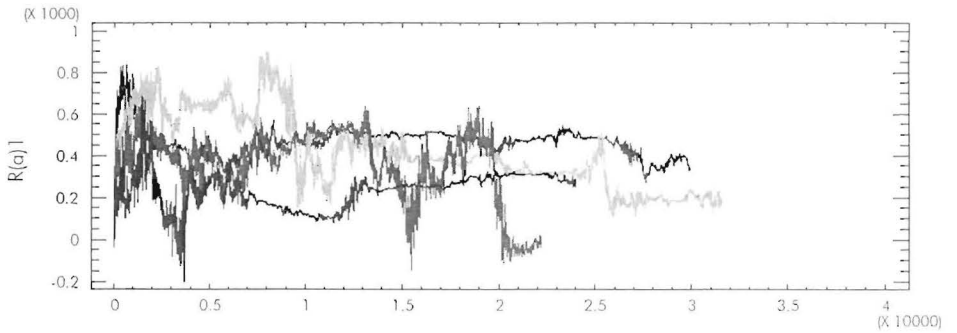


Figura 1. Trazas superpuestas de las diferentes cadenas para la respuesta correlacionada a la selección en el parámetro a de la curva de Gompertz

valores iniciales de los parámetros, recorriendo en ocasiones el mismo espacio aunque en momentos distintos. La alta correlación entre las muestras hace difícil el diagnóstico de convergencia de cada cadena considerada individualmente.

En el cuadro 1 se muestran la longitud estimada del periodo de calentamiento según el procedimiento de Raftery y Lewis, la media posterior, la desviación típica posterior y el error de Monte Carlo para la respuesta directa y correlacionada a la selec-

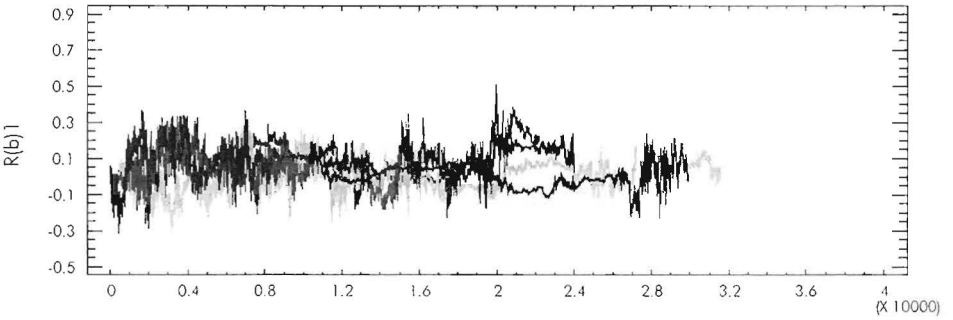


Figura 2. Trazas superpuestas de las diferentes cadenas para la respuesta correlacionada a la selección en el parámetro b de la curva de Gompertz

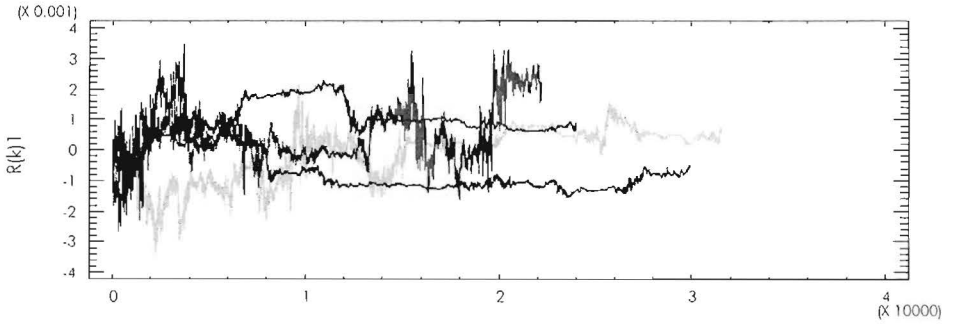


Figura 3. Trazas superpuestas de las diferentes cadenas para la respuesta correlacionada a la selección en el parámetro k de la curva de Gompertz

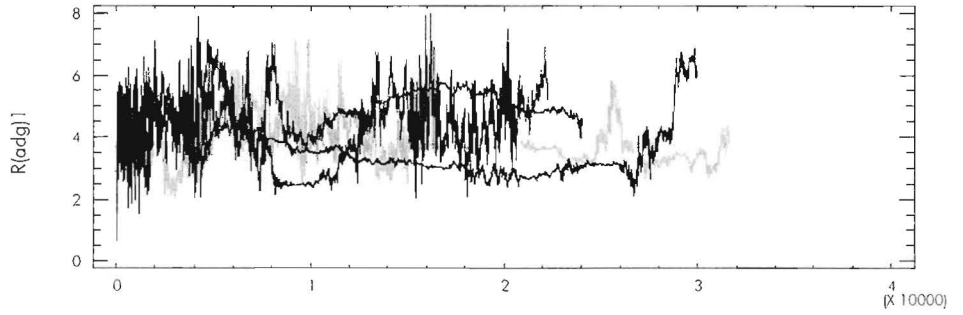


Figura 4. Trazas superpuestas de las diferentes cadenas para la respuesta directa a la selección

Cuadro 1
Longitud estimada del periodo de calentamiento (B), media posterior (PM), desviación típica posterior(SD), y error de Monte Carlo (MCSE) de la madia en cada cadena para la respuesta directa y correlacionada a la selección

	B	PM	sd	MCse	ESS
Chain 1					
R(a)	234	234	144	53,4	7,23
R(b)	126	0,0682	0,0934	0,0293	10,1
R(k)	108	0,000645	0,000718	0,000237	9,16
R(adg)	120	3,86	0,692	0,164	17,7
Chain 2					
B	324	318	60	31,1	3,74
R(a)	190	0,00917	0,0551	0,0137	16,22
R(b)	340	0,0000107	0,000321	0,000115	7,74
R(adg)	430	3,04	0,343	0,0595	33,2
Chain 3					
R(a)	1176	358	46	25,4	3,22
R(b)	360	0,0242	0,0503	0,0130	15,0
R(k)	348	-0,000715	0,000185	0,0000507	13,4
R(adg)	368	2,78	0,656	0,190	11,9
Chain 4					
R(a)	220	185	17	7,37	5,29
R(b)	153	0,0697	0,0667	0,0396	2,84
R(k)	392	0,000896	0,0000788	0,0000230	11,8
R(adg)	220	4,21	0,206	0,136	2,28

R(a), R(b), R(k) = respuesta correlacionada a la selección en los parámetros a,b,k de la curva de Gompertz.
R(dg) = respuesta directa a la selección por incremento en la ganancia media diaria (adg).
PM = Media; sd = desviación típica; P > 0 = probabilidad de una respuesta posititiva.

ción en los parámetros de la curva de Gompertz en cada cadena. El número de muestras a descartar varió entre 108 y 1176 correspondiendo el valor más elevado a la respuesta correlacionada en el parámetro *a*. De los resultados obtenidos y de la inspección visual de las cadenas se decidió utilizar las 20.000 últimas muestras de las cadenas más largas y las 10.000 últimas de las cadenas más cortas para estimar directamente de ellas aspectos descriptivos de las distribuciones marginales posteriores. La elevada correlación entre las muestras determina el elevado error de Monte Carlo y el reducido

tamaño efectivo de la muestra encontrados en cada una de las cadenas. Las diferencias entre las medias posteriores podían ser explicadas por el error de Monte Carlo por lo que las muestras en cada una de las cadenas podían ser consideradas muestras procedentes de la misma distribución estacionaria. Al aplicar el test de Gelman y Rubin para el diagnóstico de convergencia el valor más elevado correspondía a la respuesta correlacionada en el parámetro k de la curva de Gompertz para el que se obtiene un valor del factor de escala (shrink factor) de 2,58, mientras que para la mayor parte de los parámetros del modelo los valores obtenidos se encontraban cerca de 1. Se decidió pues reunir las muestras retenidas tras el periodo de calentamiento de todas las cadenas para estimar los parámetros descriptivos de las distribuciones marginales posteriores. No obstante, dado que el tamaño efectivo de las muestras era muy reducido consideramos que los resultados deben ser tomados como una primera aproximación al problema o simplemente como una aplicación numérica del procedimiento de análisis estadístico, no pudiéndose realizar inferencia acerca de la forma de las distribuciones marginales posteriores. Son necesarias más muestras para un análisis completo, sin embargo debido a la complejidad del modelo el tiempo de computo requerido sería excesivamente elevado (el análisis actual ocupó 12 días en una máquina HP exemplar S-class computer (HPPA risk 8000, 8 processors) con 280 MHz) por lo que parece necesario estudiar una alternativa implementación del proceso de muestreo que permita reducir la autocorrelación entre las muestras.

El cuadro 2 muestra algunos parámetros descriptivos correspondientes a las distribuciones marginales posteriores de la respuesta directa y correlacionada a la selección. La selección por velocidad de crecimiento ha tenido éxito lográndose un incremento de 0,51 g/d por generación lo que supone un aproximadamente un 2% sobre la media anual. Esto hace posible la reducción del periodo de cebo ya que el peso de sacrificio se mantiene constante puesto que responde a las demandas del mercado. La estima de la respuesta obtenida aquí es similar a la obtenida por ESTANY *et al.* (1992) en las dos líneas de conejos que dieron lugar posteriormente a la línea utilizada en este experimento (0,65 g/d y 0,52 g/d respectivamente) y por ROCHAMBEAU *et al.* (1989, 0,85g/d). Se observa también un incremento correlacio-

Cuadro 2
Parámetro s descriptivos de la distribución marginal posterior estimada de la respuesta directa y correlacionada a la selección

Variable	PM	sd	P > 0
R(a)	295	97	0,996
R(b)	0,0341	0,0685	0,708
R(k)	0,0000219	0,000711	0,634
R(adg)	3,28	0,757	1,000

R(a), R(b), R(k) = respuesta correlacionada a la selección en los parámetros a,b,k de la curva de Gompertz.
R(dg) = respuesta directa a la selección por incremento en la ganancia media diaria (adg).
PM = Media; sd = desviación típica; P > 0= probabilidad de una respuesta posititiva.

nado en el parámetro a, el cual describe el peso en el estado adulto, mientras que los demás parámetros de la curva de Gompertz, relacionados con la pendiente de la curva o con el peso al inicio del crecimiento, no han sido apenas afectados por la selección. Los resultados son consistentes con los obtenidos mediante la comparación entre los grupos contemporáneos, Control y Seleccionado, en análisis anteriores. Debido a que el intervalo generacional en conejo es relativamente corto (6-7 meses), las consecuencias de esto podrían ser la obtención de animales de gran tamaño, difíciles de manejar y susceptibles de sufrir determinadas patologías como el "mal de patas". El desarrollo actual de la inseminación artificial cuyo uso esta cada vez más generalizado permite reducir la cantidad de machos terminales necesarios en una granja comercial, lo cual resta importancia a algunos de los problemas que podrían plantearse.

Las estimas de las heredabilidades y correlaciones genéticas y residuales de los parámetros de la curva de crecimiento y la ganancia media diaria resultaron muy imprecisas como era de esperar dada la reducida cantidad de datos disponibles y la ausencia de información previa por lo que no se ofrecen aquí. La estima de la heredabilidad de la ganancia media diaria (PM = 0,12, sd = 0,03) coincide con el valor obtenido en un análisis univariante y con la obtenida por ESTANY *et al.* (1992) en su experimento con animales del mismo origen genético. Las estimas publicadas varían entre 0,13 y 0,48 (ROCHAMBEAU *et al.*, 1989; MOURA *et al.*, 1991; Estany *et al.*, 1992; FERRAZ y ELER, 1994; KROGMEIER *et al.*, 1994; LUKEFAHR *et al.*, 1996; MCNITT y LUKEFAHR, 1996, MOURA *et al.*, 1997). La variabilidad en las estimas podría deberse a la variedad de razas, condiciones ambientales, periodo de tiempo considerado y metodología empleada para el análisis. Las estimas más altas podrían ser parcialmente atribuidas al hecho de que la velocidad de crecimiento fuese evaluada en una más tardía en la vida del animal que en otros estudios y por tanto estuviese menos afectada por los efectos ambientales de tipo materno y de camada común. La estima de los efectos ambientales de camada de origen también coincidía con la obtenida en el análisis univariante (0,36 vs 0,37).

Consideraciones generales

Se ha implementado un procedimiento bayesiano para estimar la respuesta correlacionada a la selección por incremento en la ganancia media diaria durante el periodo de cebo, en los parámetros que definen la trayectoria individual de crecimiento de los animales como un caso particular de la metodología general descrita en GIANOLA *et al.* (1999). Este procedimiento permite, de una manera sencilla, evaluar correctamente la respuesta correlacionada a la selección en los parámetros de la curva y estimar al mismo tiempo las (co)varianzas genéticas y residuales entre ellos y el criterio de selección en situaciones en las que la estructura de datos es incompleta. El procedimiento está basado en la utilización de un modelo jerárquico no lineal en el que la descomposición de la matriz de (co)varianzas residuales del segundo estrato del modelo jerárquico en un vector de coeficientes de regresión de los parámetros de la curva sobre el criterio de selección, una matriz condicional de (co)varianzas residuales entre los parámetros de la curva dado el criterio de selección y la varianza residual del criterio de selección permite una parametrización alternativa del problema en función de estos componentes especialmente adecuada para el caso en el que gran parte de los animales no dispongan de los registros del peso a distintas edades, como ocurre habitualmente

en los experimentos que conciernen al estudio del crecimiento en poblaciones seleccionadas por otros caracteres.

Se ha asumido independencia entre datos consecutivos correspondientes a un determinado animal. Sin embargo el procedimiento permite la inclusión de un modelo autorregresivo que permita tener en cuenta estas relaciones, las cuales afectan a la estructura de los errores, y también la inclusión en el primer estrato del modelo de algunos efectos ambientales que puedan afectar el registro en un día de control lo cual podría reducir la variación residual (REKAYA, 1997). Por otra parte, podría ser interesante evaluar el efecto de la selección en algunos otros caracteres los cuales son función de los parámetros de la curva de Gompertz, tal como el grado de madurez al sacrificio, máxima velocidad de crecimiento o edad en el punto de inflexión. Esto puese realizarse mediante la implementación de un modelo secundario en el segundo estrato tal como ha sido hecho por REKAYA (1997), CHANG (1999) y VARONA *et al.* (1999) en curvas de lactación. En resumen, el procedimiento presenta una gran flexibilidad permitiendo la utilización de una gran variedad de modelos de diferente complejidad. Sin embargo pensamos que es necesaria una mayor investigación con objeto de encontrar algún procedimiento de muestreo que permita reducir las correlaciones entre las muestras, haciendo así al procedimiento menos demandante en tiempo de cómputo.

Bibliografía

CHANG Y., 1998. Bayesian Analysis of Lactation Curves in Dairy Sheep. Master of Science. Uni-versity of Wisconsin- Madison, USA..

ESTANY J., CAMACHO J., BASELGA M., BLASCO A., 1992. Selection response of growth rate in rabbits for meat production. Genetic Selection Evolution. 24: 527-537.

FERRAZ J.B.S., ELER P., 1994. Use of different animal models in prediction of genetic parameters of 23 traits of Californian and New Zealand White rabbits raised in tropics and suggestion of selection criteria. Pr. 5th World Cong.G.Appl.Liv.P. 20:348

FIORELLO C.V., GERMAN R.Z., 1997. Heterochrony within species : craniofacial growth in giant, standard, and dwaf rabbits. Evolution. 51:250-261.

GELMAN A., RUBIN D.B., 1992. Inference from iterative simulation using multiple sequences. Statistical Science. 7:457-72.

GIANOLA D., FERNANDO R.L., 1986. Bayesian methods in animal breeding theory. Journal of Ani-mal Sciences. 63:217-244.

GIANOLA D., PILES M., BLASCO A., 1999. Bayesian Inference about Parameters of a Longitudinal Trajector when Selection operates on a Correlated Trait. Proceeding of the VII International Symposium on Animal Breeding and Genetics. Viçosa. Brasil. 101-132.

GOMEZ E.A., BLASCO A., 1992. Growth curves of lines selected on growth rate or litter size. 5 th World Rabbit Congress. Oregon. B: 872-878.

KROGMEIER D., DZAPO V., MAO L., 1994. Additive genetic and maternal effects on postweaning growth and carcass traits in rabbits. Journal of Animal Breeding and Genetics. 111: 289-297.

LUKEFAHR S.D., ODI H.B., ATAKORA J.K.A., 1996. Mass selection for 70-day body weight in rabbits. Journal Animal Science. 74:1481

MCNITT J.I., LUKEFAHR S.D., 1996. Genetic and environmental parameters for postweaning growth traits of rabbits using an animal model.Proceeding of the VI World Rabbit Congress. :325.World Rabbit Science Association. Tolouse.France.

MOURA A.S.A.M.T., KAPS M., VOGT D.W., LAMBERSON W.R., 1997. Two-way selection for daily gain and feed conversion in a composite rabbit population. Journal of Animal Sciences 75:2344-2349.

MOURA A.S.A.M.T., POLASTRE R., CARMELO M.J., 1991. Genetic study of individual performance from weaning to slaughter in selecta rabbits. Journal of Applied Rabbit Re-search.14(4): 228.

RAFTERY A.E., LEWIS S., 1992. How many iterations in the Gibbs sampler? In Bayesian Statistics 4. J.O. Ber-

ger, J.M. Bernardo, A.P. Dawid and A.F.M. Smith (Eds). 763-773. Oxford University Press, Oxford.

REKAYA R., 1997. Análisis Bayesiano de datos de producción en los días del control para la selección de caracteres lecheros. Tesis Doctoral. Escuela Técnica Superior de Ingenieros Agrónomos. Universidad Politécnica de Madrid, Madrid, España.

ROCHAMBEAU H., de FUENTE L.F. de la ROUVIER R., 1989. Sélection sur la vitesse de croissance post-sevrage chez le lapin.Genetic Selection Evolution.21:527-546.

SEELAND G., ROSSLER B., RODER B., 1996. Analysis of growth in rabbits of different breeds with selected growth functions. Archiv fur Tierzucht. 39: 533-544.

VARONA L., MORENO C. GARCÍA CORTÉS L.A., ALTARRIBA J., 1997. Multiple Trait genetic analysis of underlying biological variables of production functions. Lives. Prod. Sci. 47:201-209.

VARONA L., MORENO C., GARCÍA CORTÉS L.A., ALTARRIBA J., 1998. Bayesian analysis of the Wood's lactation curve for Spanish dairy cows. J.Dairy Sci.81:1469-1478.

ITEA (2000). Vol. **96A N.° 3,** 349-361

ADVANCES IN PIG MOLECULAR GENETICS, GENE MAPPING AND GENOMICS

Max F. Rothschild

C.F. Curtiss Distinguished Professor of Agriculture, USDA Pig Genome Coordinator, Department of Animal Science, Iowa State University, Ames, Iowa 50011, USA

SUMMARY

Advances in the fields of molecular genetics and genomics have been considerable and these have led to the development of very useful linkage and physical maps of the pig genome. The genetic linkage map now has over 1,900 loci including approximately 300 genes. The physical genetic map has over 1,000 loci including a growing number of known genes. The development of new maps including those incorporating large numbers of AFLP markers map are also well underway and will add 2,000 additional markers, bringing the total to nearly 4,000 genes and markers. Several recent quantitative trait loci scans and candidate gene analyses have identified important chromosomal regions and individual genes associated with traits of economic interest in the pig. These include QTL for growth and backfat, meat quality traits, and reproduction. The causative mutations for porcine stress syndrome (HAL or CRC1) and the RN disorder, candidate genes for litter size (ESR, PRLR, RBP4), growth (MC4R), meat quality (hFABP, aFABP), disease resistance (FUT1, SLA, NRAMP), and coat color (KIT, MC1R) have also been identified. The commercial pig industry is actively using this information and in-house research to improve pig production by marker assisted selection (MAS). Furthermore, research to study the co expression of hundreds of genes is now beginning. This research will aid in our understanding of genetic systems and how to alter their relationships to improve pig production.

Introduction

Molecular genomic analysis has revolutionized how animal geneticists examine the genetic differences that exist within commercial and exotic pigs. In the past few years, efforts have been directed toward the development of genomic maps consisting of anonymous genetic markers and known genes. In addition, comparative genome maps have aided greatly in our search for interesting and potentially useful genes in the pig. The coverage on these genetic maps is now sufficient to allow researchers to conduct quantitative trait loci (QTL) linkage analyses. These QTL linkage analyses involve employing a genomic scan where generally F2 or backcross families are used and genotypes are obtained for many (>100) markers evenly spaced across the genome. Several such experiments are underway or recently completed and are beginning to produce interesting and useful results. Candidate gene and comparative mapping approaches have also been successful in identifying major genes affecting several traits. Candidate gene analyses is

undertaken when a gene is chosen based on the physiology of the trait. The candidate gene is assumed to affect trait performance. This is supplemented by comparative gene analysis that allows researchers to find "positional candidate genes" in the regions associated with possible QTL. To date, several major genes have been found with the candidate gene approach. Present and future genetic improvements will result from the more detailed genetic maps and our growing understanding of the function and structure of the individual genes and gene families that are responsible for the economically important traits in pigs. The purpose of this paper is to review the recent discoveries of gene mapping and the status of QTL and candidate gene analyses in the pig and to forecast future developments and their applications in industry.

Results

Results will be presented for status of the pig genome maps and QTL and candidate gene analyses in terms of growth and performance, meat quality, disease resistance and reproduction traits. Many of the results presented earlier (ROTHSCHILD and PLASTOW, 1999) are now updated. The author apologizes for data or reports inadvertently left out.

Status of the genome maps

The effort required to generate mapping information is very significant. Scientists and administrators realized that a cooperative effort for gene mapping in livestock would be the most productive approach. The PiGMaP project (ARCHIBALD *et al.*, 1995) was initiated in Europe and was funded by the European Economic Community. PiGMaP involved 18 European labs and a total of 7 other labs from the U.S., Japan and Australia. In the U. S., the USDA launched two efforts. The USDA-ARS directs a sizeable gene mapping project at the Meat Animal Research Center in Clay Center, Nebraska. Secondly, the National Animal Genome Research Program was developed under the direction of USDA-CSRS in 1993. This program was designed to provide a structure which would stimulate coordination and collaboration of gene mapping in all species, including pigs. Scientists from state and private universities and federal labs cooperatively created a Swine Genome Technical Committee. The U.S. Pig Genome Coordinator activities, in concert with activities of the USDA-ARS and other international gene mapping projects have allowed the status of the pig gene map to evolve more quickly in the last several years. In 1989, only 50 genes and markers were mapped in the pig, and many of these were isozyme or blood group markers. At present, nearly 1,900 markers and genes have been mapped in pigs (ARCHIBALD *et al.*, 1994 and personal communication; MARKLUND *et al.*, 1996a; ROHRER *et al.*, 1996), though not all have been published. There are two primary maps. The most extensive pig map has about 1,100 loci (ROHRER *et al.*, 1996), consisting mostly of anonymous markers. A second PiGMaP publication is under consideration. In addition, the new PiGMaP genetic linkage map will have over 700 genes and markers on its map, including about close to 300 genes. Coverage for the different linkage maps varies, and while the average distance between markers is approximately 3-5 cM, some large gaps still remain. Efforts are underway to produce an integrated map including AFLP (Amplified Fragment Length Polymorphism) markers which is expected to have eventually 2000 AFLP markers (Archibald and collea-

gues, personal communication). Physical maps for all the species used to lag behind the genetic linkage maps but are catching up quickly due to some excellent resources such as somatic cell hybrids for physical mapping and a radiation hybrid map developed in France and being used by several groups (HAWKINS *et al.*, 1999). The physical genetic map in the pig currently consists of over 600 genes and markers, while the radiation hybrid panel map has over 1,000 markers. Despite these concerted efforts, the pig genome map still pales in comparison to the human and mouse maps. However these gene rich species provide real help to our efforts in the pig. As more genes are placed on the pig comparative map, the human genome map will have increased utility in the search for individual genes responsible for traits of economic interest.

Growth and performance traits

The first successful large QTL analysis was conducted using a Wild Boar and Large White three generation family and revealed major QTL accounting for 20% of the phenotypic variance for average backfat and abdominal fat on chromosome 4 (ANDERSSON *et al.*, 1994). A QTL for growth was also found on chromosome 13 accounting for 7% to 12% of the phenotypic variation. Additional confirmation of the chromosome 4 results have been seen in subsequent generations (MARKLUND *et al.*, 1996b) and by other experiments using Chinese pig crosses (WANG *et al.*, 1998; MILAN *et al.*, 1998; Moser and Cepica, personal communication). Using a candidate gene analysis PIT1 was identified to be associated with backfat and birth weight (YU *et al.*, 1995) and it maps in the center of the chromosome 13 QTL found by ANDERSSON *et al.* (1994). Recent results have demonstrated that in the Iowa State University Meishan x American F2 families a QTL for early growth is centered above PIT1 and a QTL for backfat is 20 cM from PIT1 (YU *et al.*, 1999). Results from another specialized meat quality family confirms such a backfat QTL in this region (MALEK *et al.*, 2000; Rothschild, unpublished results).

The pig major histocompatibility complex (MHC) spans the centromere of chromosome 7. Associations between MHC haplotypes and several traits have been reported over the years. These have been confirmed, in part, using MHC class I DNA probes (JUNG *et al.*, 1989). More recently, QTL scans have identified QTL for growth and backfat traits in Chinese crosses on chromosome 7 (ROTHSCHILD *et al.*, 1995; BIDANEL *et al.*, 1996; CHEVALET *et al.*, 1996; MOSER *et al.*, 1998; Rohrer and Keele 1998a; WANG *et al.*, 1998). The backfat and birth weight QTL centered near the region of TNFA and S0102. A candidate gene analysis involving AMPEPN, which maps to the same region, also revealed an association with growth rate (NIELSEN *et al.*, 1996). The overall results to date suggest that at least one growth and backfat QTL exists in this region. Other results have included a growth trait QTL on chromosome 6, but it seems to be associated with the known effect caused by the RYR1 (CRC1) gene causing malignant hyperthermia (GELDERMANN *et al.*, 1996) or other unknown genes in the immediate vicinity of RYR1. Some additional associations have been reported for chromosome 1 (ROHRER and KEELE, 1998a; PAZEK *et al.*, 1999), chromosome 3 (CASAS-CARRILLO *et al.*, 1997b), chromosomes 6 and 8 (WILKIE *et al.*, 1996), chromosome 14 (BIDANEL *et al.*, 1996) and the X chromosome (ROHRER and KEELE, 1998a). New data that is soon to be published shows signifi-

cant QTL for growth on additional chromosomes (ROTHSCHILD *et al.*, in preparation).

A combination of QTL scans and candidate gene analyses using GH have been performed for chromosome 12 for several performance measures (NIELSEN *et al.*, 1995; CASAS-CARRILLO *et al.*, 1997a; KNORR *et al.*, 1997) but results were not in general agreement. Analysis of the chromosome 5 region near IGF-1 revealed significant effects on average daily gain (CASAS-CARRILLO *et al.*, 1997a). Additionally, GERBENS *et al.* (1997) and TEPAS *et al.* (1996) reported associations of the heart fatty acid-binding protein (H-FABP) gene with intramuscular fat and the myogenin gene with average daily gain. Other candidate genes including CCK and CCKAR (CLUTTER *et al.*, 1996, leptin (reference to be added) and the leptin receptor (VINCENT *et al.*, 1997) have been mapped and may prove to be associated with growth, fatness and appetite traits.

Perhaps the most interesting new results are those dealing with melanocortin 4 receptor (MC4R) and its association with feed intake, growth and backfat (KIM *et al.*, 2000). Using information from a mouse knockout experiment done by Millennium Pharmaceuticals, the Iowa State University group in collaboration with PIC found a naturally occurring missense mutation that causes the pig to eat more (about 10%), grow faster (6-8%), and grow fatter (6-10%). This mutation can be used to improve feed intake in certain sow lines where fat is already at the desired level and in sire lines to reduce the level of fat. This new genetic test is patented and being used in the industry.

Meat quality traits

The earliest QTL research in meats was by Andersson and colleagues (ANDERSSON-EKLUND *et al.*, 1996). They have conducted one of the most complete QTL scans for meat quality using 234 markers on 191 F2 animals. QTL for several meat quality traits (pH, water holding capacity and pigmentation) were found to be on chromosome 2 and chromosome 12. Our earliest work at Iowa State University demonstrated some association of meat color and firmness scores with regions on chromosomes 4 and 7 (ROTHSCHILD *et al.*, 1995, WANG *et al.*, 1998). Additional associations with meat quality traits have been reported for chromosome 7 (MOSER *et al.*, 1998) and for number of muscle fibers on chromosome 3 (MILAN *et al.*, 1998). Two other interesting associations related to meat quality have been noted for chromosome 7. The activity of Malic enzyme, a lipogenic enzyme in muscle has been shown to be associated with the SLA complex on chromosome 7 (RENARD *et al.*, 1996). Furthermore, a directed QTL scan revealed that there was a major QTL for androstenone level which is associated with boar taint in the region of the SLA complex (BIDANEL *et al.*, 1996; MILAN *et al.*, 1998). Additional QTL results have also been reported for carcass composition and wholesale product yield traits. ROHRER and KEELE (1998b) reported QTL for loin eye area (chromosome 1), weight of trimmed ham, loin, picnic and Boston butt (chromosomes 1 and X) and carcass length (chromosome 7).

Most recently, Rothschild and colleagues conducted a major meat quality QTL scan using two commercial breeds, Berkshire and Yorkshire. The F2 population consisted of 525 animals that were genotyped for 125 markers and measured for 40 traits for growth, carcass, meat and sensory qualities. Rothschild and colleagues (MALEK *et al.*, 2000) found over 100 significant QTL at the 5% chromosome significance level and nearly 20 at the 5% genome significance level.

Evidence for a major gene for intramuscular fat (IMF) was obtained using segregation analysis (JANSS *et al.*, 1994, 1997), and this group went on to try to identify the gene using QTL mapping. Although they recently reported evidence for suggestive linkages with markers on chromosome 1 and a marker on chromosome 3, they were not able to find identify the major gene suggested by the segregation analysis.

It has been known for some time that pale soft and exudative (PSE) pork is associated with variation in the RYR1 gene on chromosome 6. This has been well demonstrated in a QTL scan for several meat quality traits related to PSE in an F2 population originating from a Pietrain background (GELDERMANN *et al.*, 1996). Focus has also centered on a gene originating in Hampshires called the RN gene that is associated with lower pH and increased glycogen content in the meat. The RN gene was first mapped to chromosome 15 (MILAN *et al.*, 1996; MARIANI *et al.*, 1996a; REINSCH *et al.*, 1997) and surrounded by flanking markers. The actual gene had alluded researchers for some years, but DNA tests using markers for the RN gene were utilised by researchers (DE VRIES *et al.*, 1997). Early this year a consortium led by Leif Andersson, Denis Milan and Christian Looft reported the identification of the causative mutation. The gene involved is a new member of a gene family coding AMP-activated protein kinases (named PRKA). Interestingly, the same gene might explain certain forms of diabetes in humans and the consortium is looking at the opportunities for their research to benefit human health. This result represented a tremendous effort by this group of labs to move from the chromosomal position to the gene itself. This new gene discovery result represents another important addition to be used by pig breeders to improve meat quality where the Hampshire breed is used.

Among candidate genes investigated for muscle quality is the H-FABP gene that may be associated with intramuscular fat (GERBENS *et al.*, 1997). These workers identified polymorphisms in this gene and found them to be associated with variation in IMF in the Duroc (GERBENS, 1998). The H-FABP gene maps to pig chromosome 6 in an area where some recent reports suggest a QTL in the region. A comparison of the homozygous haploid classes found that they differed by about 15% of the mean value. Interestingly, the difference in IMF content is only partially explained by backfat content. This should permit selection for increased IMF based on H-FABP genotype as long as increased backfat is countered by ongoing phenotypic backfat selection. It will be interesting to see how effective and practical this test will be in the near future. This test is patented and is available through license arrangements (MERKS, personal. comm.).

Other markers which have been generated for meat quality based on the candidate gene approach include myogenin (increased fibre number, which may impact overall pork quality) (SOUMILLION *et al.*, 1997) and calpastatin (ERNST *et al.*, 1997). Coat color, though not directly associated with meat quality is of interest to the packing industry. White pigs are preferred at slaughter and the cost of a "colored carcass" at some locations may be over $1/pig. Andersson and colleagues (JOHANSSON-MOLLER *et al.*, 1996) have now identified the KIT gene as that responsible for white coat color and a DNA test is patented and being used in MAS programs (ANDERSSON and PLASTOW, personal communication). The MC1R gene has also been identified to control red and black color in the pig (MARIANI *et al.*, 1996b) and the gene test is likely to be made public in the U.S. this year.

While the normal inheritance of genetic factors is expected sometimes imprinting exists. The clearest example in pigs is the effect of the IGF2 region of chromosome 2 on muscling in pigs of different backgrounds (JEON *et al.*, 1999, NEZER *et al.*, 1999). The ability to predict the genotype at such loci is a clear benefit of MAS.

Reproduction traits

Given the necessity of larger resource families and the difficulty and time required to obtain information on reproduction traits, it is not surprising that results of QTL scans for these traits are limited. Initial scans have revealed promising results on chromosome 8. WILKIE *et al.* (1996) reported possible QTL for uterine length and ovulation rate, though in different chromosomal positions. RATHJE *et al.* (1997) reported a sizable QTL for ovulation rate (+3.07 ova) on chromosome 8 also but some distance from the ovulation QTL observed by Wilkie and colleagues. Later work in the same lab did not confirm this finding (POMP and JOHNSON, personal communication). In the French QTL experiment (MILAN *et al.*, 1998) a QTL for increased litter size of one piglet was found in the same location on chromosome 8 as Rathje. The large ovulation rate/litter size QTL on chromosome 8 is of interest as it mapped to the region which is synthetic to the Booroola fecundity gene in sheep. Interestingly, using a single microsatellite marker (OPN) in the same chromosome 8 region, (SHORT *et al.*, 1997b; VAN der STEEN *et al.*, 1997) also found significant effects for litter size in commercial lines. Further evidence exists for at least one QTL for litter size components at the top of chromosome 8 (ROHRER *et al.*, 1999) and perhaps another nearer the centromere (Plastow, Personal communication). Limited chromosome QTL analyses have further suggested other reproductive QTL on chromosomes 4 and 6 (WILKIE *et al.*, 1996), on chromosome 7 (WILKIE *et al.*, 1996; MILAN *et al.*, 1998) and chromosomes 4, 13, 15 (RATHJE *et al.*, 1997).

Candidate gene analysis for reproduction has shown considerable merit. Results have clearly demonstrated that the estrogen receptor (ESR) is significantly associated with litter size (ROTHSCHILD *et al.*, 1996; SHORT *et al.*, 1997a). Estimates of allelic effects vary from 1,15 pig/litter in Meishan synthetics to .42 pigs/litter in Large White lines. These results have not been confirmed by QTL scans using divergent crosses involving Meishan and Large White pigs. This may have resulted from small sample sizes or the fact that the ESR gene allele was not segregating in the populations involved in the QTL scans. The ESR marker was incorporated successfully into the PIC selection indices for Large White based dam lines, resulting in an increase in the rate of genetic response in its nucleus herds (SHORT, WILSON, MCLAREN and PLASTOW, unpublished results). Furthermore, the increase in average litter size is observed in crossbred products derived from these lines.

Other effects have been reported for retinoic acid receptor gamma (RARG), melatonin receptor 1A (MTNR1A) (MESSER *et al.*, 1996; OLLIVIER *et al.*, 1997) and follicle stimulating hormone beta (FSHB) genes (LI *et al.*, 1998). Researchers at Iowa State University, working with PIC have demonstrated that the prolactin receptor (PRLR) locus is significantly associated with litter size (VINCENT *et al.*, 1998). This has been confirmed in two smaller studies (Ohio State Univ. and WAU in the Netherlands) now being submitted for publication. More recently, retinol binding protein 4 (RBP4) has been shown in a study involving nearly 2,500 litters to be associated with an increase of about 0.25

pigs per litter (ROTHSCHILD *et al.*, 2000). Most of the candidate gene analyses have involved considerably more sows and litters than the QTL analyses and this might explain the lack of QTL scan confirmation of the regions in which the candidate gene effects have been reported to date.

Disease resistance and immune response traits

To date, QTL scans for disease resistance or immune response QTL have been limited. An exception is work underway by Andersson and colleagues (EDFORS-LILJA *et al.*, 1998) to study some immune response parameters. Some immune capacity QTL have been identified. Also, a QTL for base cortisol level which may be related to stress and perhaps immune response, has been mapped to the end of chromosome 7 (MILAN *et al.*, 1998). Analyses of associations of candidate genes have been more substantial. The general location of the K88 E. coli receptor has been known for some time and fine mapping and candidate gene analysis of the region is underway in many labs. Two alpha (1,2) fucosyltransferase genes (FUT1, FUT2) on porcine chromosome 6q11 have been identified and are closely linked to the blood group inhibitor (S) and Escherichia coli F18 receptor (ECF18R) loci (MEIJERINK *et al.*, 1997). This work has now confirmed that a polymorphism in these genes is closely linked to ECF18R in Large White, Landrace, Hampshire, Duroc and Pietrain pigs and it could be a good marker for marker assisted selection of E. coli F18 adhesion resistant animals in these breeds. Clearly now the FUT1 or FUT2 gene products are involved in the synthesis of carbohydrate structures responsible for bacterial adhesion remains to be determined. The SLA complex on chromosome 7 has recently been associated with resistance to primary infections with Trichinella spiralis but not to resistance to toxoplasmosis (Lunney, personal communication). The gene for Natural Resistance Associated Macrophage Protein 1 (NRAMP1), associated with resistance to Salmonella challenge in mice, has been recently mapped to pig chromosome 15 (SUN *et al.*, 1998). Several other candidate genes are being investigated. Genes associated with human disorders, which have been identified and mapped in the pig, include clotting factor IX (SIGNER *et al.*, 1996) and the familial hypercholesterolaemia gene (HASLER-RAPACZ *et al.*, 1996).

Genetic diversity

A final area for the use of molecular markers and genes is in the area of genetic diversity. This has traditionally been accomplished by examining at first enzyme polymorphisms and more recently using microsatellites. A final example of this approach is the European biodiversity project, coordinated by L. Ollivier from France (LAVAL *et al.*, 2000). Some arguments have been placed that such approaches with anonymous markers are not sufficient to determine functional diversity and that polymorphisms within real genes should be used (CIOBANU *et al.*, 1999; Rothschild, 1999). This approach was used by Ciobanu and colleagues (CIOBANU *et al.*, 1999) and demonstrates differences in functional genes between two native pig breeds of Romania.

Potential of DNA marker assisted selection in the pig industry

Information at DNA level can help industry breeders and geneticists to fix a specific major mutation, such as the normal

Halothane allele. It can also be used to assist in the selection of quantitative traits including those that can be selected by traditional means (see table 1). Molecular information can increase the accuracy of selection and therefore the selection response. The size of the extra response obtained through Marker Assisted Selection (MAS) schemes has been considered by many workers from a theoretical point of view. GIBSON (1994) and others have shown that there is a short term benefit in using MAS, but that in some cases this can lead to a long term penalty. However, this is over a relatively long time frame. MEUWISSEN and GODDARD (1996) considered a different set of assumptions and in particular looked at the impact for traits such as reproduction and meat quality that are difficult to progress using traditional methods. DEKKERS and van ARENDONK (1998) have shown that these results can be further improved by using control theory to optimise response. Altogether these results are extremely encouraging. Importantly, these responses can be sustained if new markers are continually identified. For example, new markers can be added to the selection index as old markers begin to reach fixation.

In the meantime we anticipate that significant progress will be made by utilizing candidate genes and searching for population-wide linkage disequilibrium, using tools such as AFLP. It is anticipated that AFLP analysis will be an extremely powerful tool as it generates large numbers of markers rapidly in interesting populations. AFLP can be used with bulk segregant analysis (BSA) in a form of map-less QTL analysis. For example, markers were identified close to the dominant white locus (PLASTOW *et al.*, 1998). Another tool will be the use of large numbers of SNPs (single nucleotide polymorphisms) and also making use of for population-wide linkage disequilibrium.

Discussion and future developments

Additional developments in the genome maps are expected but the generation of hundreds of new random microsatellite markers is not likely. The addition of an AFLP map is likely to aid greatly in finding QTL. In addition to more extensive scans of existing populations, many new QTL experi-

Table 1
Molecular Genetic Tests Used by the Swine Industry

Parentage tests	Non exclusive use
HAL	Meat quality - non exclusive use
ESR	Litter size - exclusive use (PIC)
PRLR	Litter size - exclusive use (PIC)
RBP4	Litter size - exclusive use (PIC)
KIT	White color - exclusive use (PIC)
MC1R	Red/black color - exclusive use (PIC)
MC4R	Growth and fatness - exclusive use (PIC)
FUT1	Edema disease E. coli F18 - exclusive use (PIC/ITH Switzerland)
RN	Meat quality - non exclusive tests (soon) (Uppsala, INRA, Kiel).
AFABP, HFABP	Intramuscular fat - non exclusive (IPG)
IGF2	Carcass composition - exclusive use (Seghers)
Trade secret tests	Several traits - many companies

ments are underway for performance, reproduction and meat quality traits (MORAN, personal communication; NEZER *et al.* 1996; van OERS *et al.* 1996). In the short time it has taken to build a considerable genetic map, many QTL and candidate analyses have yielded interesting results. The QTL scans have identified several chromosomes that are now targets for further confirmation of the chromosomal region, advanced fine mapping of the QTL, and positional comparative candidate gene analysis. This will be aided by the ever improving comparative gene maps. Additional experiments will either confirm the regions and lead to the eventual isolation of the gene or genes of interest or will produce conflicting results. Such conflicting results may be the results of haplotype effects, epistasis or background genotype effects, or sampling. Given that we wish to detect not only large but also moderate gene effects, experimental size will need to be increased or several experimental results will need to be pooled. Recently, several PiGMaP participants pooled information on chromosome 4 to obtain additional power in the analysis (WALLING *et al.*, 2000). Additional plans are underway to use this approach to join other QTL experiments. The relevance of results found in crosses involving the Wild Boar and the Chinese breeds will also have to be evaluated in commercial lines. This is the objective of a EC funded project on the utilisation of QTL (ANDERSSON and PLASTOW, personal communication) and the work by Rothschild and colleagues (MALEK *et al.*, 2000)

New technical develops continue to be provided which may yield exciting results. Several EST (expressed sequence tags) projects are underway in the pig. These projects attempt to identify bits of new genes from specified tissues. Many of these will be mapped and the comparative map of the pig will advance rapidly. The use of DNA chips or arrays to examine gene expression will offer new glimpses into the complex traits of economic importance in the pig. These and other yet to be developed methods offer great promise for the future of swine improvement.

Acknowledgements

The author appreciates the invitation to attend this conference and gratefully acknowledges the many colleagues around the world who kindly provided information in both published and unpublished forms for inclusion in this review. Comments from and discussion with Graham Plastow are greatly appreciated. The support of the Iowa Agriculture and Home Economics Experiment Station, the USDA-CSREES Pig Genome Coordination program and the industry support of PIC Int. Group and PIC USA are appreciated. Support from colleagues in the EC PiGMaP program and those in the bioinformatics division at Roslin are also gratefully acknowledged.

References

ANDERSSON-EKLUND L., MARKLUND L., LUNDSTRÖM K., ANDEERSSON K., HANSSON I., LUNDHEIM N., MOLLER M., ELLEGREN H., ANDERSSON L., 1996. Mapping QTLs for morphological and meat quality traits in a wild boar intercross. Anim. Genet. 27 (Suppl. 2): 111.

ANDERSSON L., HALEY C.S., ELLEGREN H., KNOTT S.A., JOHANSSON M., ANDERSSON K., ANDERSSON-EKLUND L., EDFORS-LILJA I., FREDHOLM M., HANSSON I., HAKANSSON J., LUNDSTROM K., 1994. Genetic mapping of quantitative trait loci for growth and fatness in pigs. Science 263: 1771-1774.

ARCHIBALD A., BROWN J., COUPERWHITE S., MCQUEEN H., NICHOLSON D., HALEY C., COUPPIETERS W., VAN DE WEGHE A., STRATIL A., WINTERO A., FREDHODLM M., LARSON N., NIELSEN V., MILAN D., WOLOSZYN N., ROBIC A., DALENS M., RIQUET J., GELLIN J., CARITEZ J-C., HUE D., BURGAUD G., OLLIVIER L., BIDANEL J-P., VAIMAN M., RENARD C., GELDERMANN H., DAVOLI R., RUYTER D., VERSTEGE E., GROENEN M., DAVIES W., HOYHEIM B., KEISERUD A., ANDERSSON L., ELLEGREN H., JOHANSSON M., MARKLUND L., MILLER R., DEAR D., SIGNERE., JEFFREYS A., MORAN C., LE TISSIER P., MULADNO ROTHSCHILD M., TUGGLE C., VASKE D., HELM J., LIU H.C., RAHMAN A., YU T.P., LARSON R.G., SCHMITZ C., 1994. The PiGMap consortium linkage map of the pig (Sus scrofa). Mammalian Genome 6:157-175.

BIDANEL J.P., MILAN D., CHEVALET C., WOLOSZYN N., CARITEZ J.C., GRUAND J., LE ROY P., BONNEAU M., RENARD C., VAIMAN M., GELLIN J., OLLIVIER L., 1996. Chromosome 7 mapping of a quantitative trait locus for fat androstenone level in Meishan X Large White F2 entire male pigs. Proceedings EAAP meetings.

CASAS E., PRILL-ADAMS A., PRICE S.G., CLUTTER A.C., KIRKPATRICK B.W., 1997a. Relationship of growth hormone and insulin-like growth factor-1 genotype with growth and carcass traits in swine. Anim. Genet. 28:88-93.

CASAS E., PRILL-ADAMS A., PRICE S.G., CLUTTER A.C., KIRKPATRICK B.W., 1997b. Mapping genomic regions associated with growth rate in pigs. J. Anim. Sci. 75:2047-2053.

CHEVALET C., BIDANEL J.P., LE ROY P., MILAN D., VAIMAN M., RENARD C., WOLOSZYN N., BOURGEOIS F., CARRITEZ J.C., GRUAND J., GELLIN J., OLLIVIER L., 1996. Mapping quantitative trait loci in a Meshian X Large White F2 population. Anim. Genet. 27 (Suppl. 2):112.

CIOBANU D.C., NAGY A. WALES R., DAY A.E., PLASTOW G.S., 1999. Genetic identification of two local Romanian pig breeds assessed using DNA markers, Proc. 50th Annual Meeting of the EAAP, Zurich, 22-26 August.

CLUTTER A.C., POMP D., SASAKI S., 1997. Development of polymorphic markers within candidate loci for appetite regulation and estimation of allele frequencies in lines of pigs with divergent genetic potential for feed intake. Anim. Genet. 27 (Suppl. 2):112.

DEKKERS J.C.M., VAN ARENDONK J.A.M., 1998. Optimizing selection for quantitative traits with information on an identified locus in outbred population. Genetical Research 71:257-275.

DE VRIES A.G., TIMM H.H., WILSON E.R., EVANS G., KELLER V., PLASTOW G.S., 1997. An effective DNA-marker test for the "acid meat" problem. Proceedings EAAP meetings, pág 340.

EDFORS-LILJA I, WATTRANG E, MARKLUND L, MOLLER M, ANDERSSON-EKLUND L, ANDERSSON L, FOSSUM C., 1998. Mapping quantitative trait loci for immune capacity in the pig. J. Immunol. 161(2):829-835.

ERNST C. W., ROBIC A., YERLE M., WANG L., ROTHSCHILD M.F., 1998. Mapping of calpastatin and three microsatellites to porcine chromosome 2q2.1-q2.4. Anim. Genet. 29:212-215.

GELDERMANN H., MULLER E., BEECKMANN P., KNORR C., YUE G., MOSER G., 1996. Mapping of quantitative trait loci by means of marker genes in F2 generations of wild boar Piétrain and Meishan pigs. J. Anim. Breed. Genet. 113:381-387.

GERBENS F., RETTENBERGER G., LENSTRA J.A., VEERKAMP J.H., TEPAS M.F.W., 1997. Characterization, chromosomal localization, and genetic variation of the porcine heart fatty acid-binding protein gene. Mammal. Genome. 8:328-332.

GERBENS F., JANSEN A., VAN ERP A.J.M., HARDERS F., MEUWISSEN T.H.E., RETTENBERGER G., VEERKAMP J.H., TE PAS M.F.W., 1998. The adipocyte fatty acid-binding protein locus: characterization and association with intramuscular fat content in pigs. Mammalian Genome 9: 1022-1026.

HASLER-RAPACZ J., ELLEGREN H., ANDERSSON L., KIRKPATRICK B.W., KIRK S., FRIDOLFSSON A.K., RAPACZ J., 1996. Mapping of a major locus for familial hypercholesterolaemia to porcine chromosome 2 and identification of a candidate gene; the low-density lipoprotein receptor. Anim. Genet. 27 (Suppl. 2): 113.

HAWKINS R.J., MURTAUGH J., FLICKINGER G.H., YERLE M., ROBIC A., MILAN D., GELLIN J. BEATTIE C.W., SCHOOK L.B., ALEXANDER L.B., 1999. A first generation porcine whole genome radiation hybrid map. Mammal. Genome: 10: 824-830.

JANSS L.L.G., VAN ARENDONK J.A.M., E.W. BRASCAMP E.W., 1997. Segregation analyes for presence of major genes affecting growth, backfat and litter size in Dutch Meishan crossbreds. J. Anim. Sci. 75: 2864-2876.

JEON J-T., CARLBORG O., TORNSTEN A., GIUFFRA E., AMARGER V., CHARDON P., ANDERSSON-EKLUND L., ANDERSSON K., HANSSON I., LUNSTROM K., ANDERSSON L., 1999. A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus. Nature Genetics 21: 157-158.

JOHANSSON MOLLER M., CHAUDHARY R., HELLMÉN E., HÖYHEIM B., CHOWDHARY B., ANDERSSON L., 1996. Pigs with the dominant white coat color phenotype carry a duplicate of the KIT gene encoding the mast / stem cell growth receptor. Mammal. Genome 7: 822-830.

JUNG Y.C., ROTHSCHILD M.F., FLANAGAN M.P., CHRISTIAN L.L., WARNER C.M., 1989. Association of restriction fragment length polymorphisims of swine leucocyte class I genes with production traits of Duroc and Hampshire boars. Anim. Genet. 20: 79-91.

KIM K.S., LARSEN N., SHORT T., PLASTOW G., ROTHSCHILD M.F., 2000. A missense variant of the melanocortin 4 receptor (MC4R) gene is associated with fatness, growth and feed intake traits. Mammal. Genom. 11: 131-135.

KNORR C., MOSER G., MÜLLER E., GELDERMANN H., 1997. Associations between GH gene variations with performance traits in F2 generations of European Wild Boar, Piétrain and Meishan Pigs. Anim. Genet.28: 124-128.

LI N., ZHAO Y.F., XIAO L., ZHANG F.J., CHEN Y.Z., DAI R.J., ZHANG J.S., SHEN S.Q., CHEN Y.F., WU C.X., 1998. Candidate gene approach for identification of genetic loci controlling litter size in swine. Proc 6th World Congress Quatitative Genetics of Livestock 26: 183-186.

MALEK M., DEKKERS J.C.M., LEE H.K., BAAS T.J., PRUSA K., HUFF-LONERGAN E., ROTHSCHILD M.F., 2000. A quantitative trait loci analysis for growth and meat quality traits in the pig. EAAP Proceedings (in press).

MARIANI P., LUNDSTRÖM K., GUSTAFSSON U., ENFÄLT A.C., JUNEJA R.K., ANDERSSON L., 1996a. A major locus (RN) affecting muscle glycogen content is located on pig chromosome 15. Mammal. Genome 7: 52-54.

MARIANI P., MOLLER M.J., HOYHEIM B., MARKLUND L., DAVIES W., ELLEGREN H., ANDERSSON L., 1996b. The extension coat color locus and loci for blood group O and tyrosine aminotransferase are on pig chromosome 6. Journal of Heredity 87: 272-276.

MARKLUND L., JOHANSSON MOLLER M., HOYHEIM B., DAVIES W., FREDHOLM M., JUNEJA R.K., MARIANI P., COPPIETERS W., ELLEGREN H., ANDERSSON L., 1996a. A comprehensive linkage map of the pig based on a wild pig - Large White intercross. Anim. Genet. 27: 255-269.

MARKLUND L., NYSTRÖM P-E., STERN S., ANDERSSON. L., 1996b. Further characterization of a major QTL for fatness on pig chromosome 4. Anim. Genet. 27 (Suppl. 2): 114.

MESSER L., WANG, L., LEGAULT C., ROTHSCHILD M.F., 1996. Mapping and investigation of candidate genes for litter size in French Large White pigs. Anim. Genet. 27 (Suppl. 2): 114.

MEIJERINK E., FRIES R., VOGELI P., MASABANDA J., WIGGER G., STRICKER C., NEUENSCHWANDER S., BERTSCHINGER H.U., STRANZINGER G., 1997. Two alpha (1,2) fucosyltransferase genes on porcine chromosome 6q11 are closely linked to the blood group inhibitor (S) and Escherichia coli F18 receptor (ECF18R) loci. Mamm Genome. 10: 736-741.

MEUWISSEN T.H.E., GODDARD M.E., 1996. The use of marker haplotypes in animal breeding schemes. Genet. Sel. Evol., 28: 161-176.

MILAN D., WOLOSZYN N., GITEAU M., NAVAS A., YERLE M., ROGEL-GAILLARD C., CHARDON P., GELLIN J., ELSEN J.M., LEROY P., 1996. Toward the identification of RN gene involved in meat quality in pigs. Anim. Genet. 27 (Suppl. 2): 114.

MILAN D., BIDANEL J.P., LE ROY, P., CHEVALET C., WOLOSZYN N., CARITEZ J.C., GRUAND J., BONNEAU M., LAFAUCHEUR L., RENARD C., VAIMAN M., MORMÈDE P., DÉSAUTÉS C., GELLIN J., OLLIVIER L., 1998. Current status of QTL detection in Large White x Meishan crosses in France. Proc 6th World Congress Quatitative Genetics of Livestock 26: 414-417.

MILAN D., JEON J.T., LOOFT C., AMARGER V., THELANDER M., ROBIC A., ROGEL-GAILLARD C., PAUL S., IANNUCCELLI N., RASK L., RONNE H., LUNDSTRÖM K., REINSCH N., GELLIN J., KALM E., LE ROY P., CHARDON P., ANDERSSON L., 2000. A mutation in an AMP-activated protein kinase g subunit is associated with excess glycogen in pig skeletal muscle. Science (in press)

MOSER G., MULLER E., BEECKMANN P., YUE G., GELDERMANN H., 1998. Mapping QTLs in F2 generations of Wild Boar, Pietran and Meishan pigs. Proc

6th World Congress Quatitative Genetics of Livestock 26: 478-481.

NEZER C., HANSET R., MOREAU L., GROBET L., SCALAIS S., GEORGES M., 1996. Mapping of genes for growth and carcass properties in Belgian Piétrain and Réhal breeds. Anim. Genet. 27 (Suppl. 2): 115.

NEZER C., MOREAU L., BROUWERS B., COPPIETERS W., DETILLEUX J., HANSET R., KARIM L., KVASZ A., LEROY P., GEORGES M., 1999. An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. Nature Genetics 21: 155-156.

NIELSEN V.H., LARSEN N.J., AGERAARD N., 1995. Association of DNA polymorphism in the growth hormone gene with basal plasma growth hormone concentration and production traits in pigs. J. Anim. Breeding and Genet. 112: 205-21.

NIELSEN V.H., DAVOLI R., KORSGAARD I.R., THOMSEN B., BIGI D., RUSO V., 1996. Association of a DNA Polymorphism in the aminopepitdase N (ANPEP) locus with production traits in pigs. Anim. Genet. 27 (Suppl. 2): 115.

OLLIVIER L., MESSER L.A., ROTHSCHILD M.F., LEGAULT C., 1997. The use of selection experiments for detecting quantitative trait loci. Genet. Research 69: 227-232.

PASEK A., WILKIE P., FLICKEENGER G., ROHRER G., ALEXANDER L., BEATTIE C., SCHOOK L., 1999. Interval mapping of growth in divergent swine cross. Mammal. Genome. 10: 117-122.

PLASTOW G.S., KUIPER M., WALES R., ARCHIBALD A.L., HALEY C.S., SIGGENS K.W., 1998. AFLP for mapping and QTL detection in commercial pigs. Proc 6th World Congress on Genetics Applied to Livestock Production 26: 209-212.

RATHJE T.A., ROHRER G.A., JOHNSON, R.K., 1997. Evidence for quantitative trait loci affecting ovulation rate in pigs. J. Anim. Sci. 75: 1486-1494.

REINSCH N., LOOFT CHR., RUDAT I., KALM E., 1996. The Kiel RN experiment: final porcine chromosome 15 mapping results. J. Anim. Breed. Genet. 114: 133-142.

RENARD CH., MOUROT J., NUNES M., LAHBIB-MANSAIS Y., GEFFROTIN C., BOURGEAUX, N., CARRITEZ J.C., GÖTZ K., BIDANEL J.P., VAIMAN M., 1996. Association between the swine MHC region and malic enzyme activity in muscle. Animal Genetics 27 (Suppl. 2): 115

ROHRER G.A., ALEXANDER L., HU Z., SMITH T.P., KEELE J.W., BEATTIE C., 1996. A comprehensive map of the porcine genome. Genome Research 6: 371-391.

ROHRER G.A., FORD J.J., WISE T.H., VALLET J.L., CHRISTENSON R.K., 1999. Identification of quantitative trait loci affecting female reproductive traits in a multigeneration Meishan-White composite swine population. J Anim Sci. 77:1385-1391.

ROHRER G.A., KEELE J., 1998a. Identification of quantitative trait loci affecting carcass composition in swine: I. Fat deposition traits. J. Anim. Sci. 76: 2247-2254.

ROHRER G.A., KEELE J., 1998b. Identification of quantitative trait loci affecting carcass composition in swine: II. Muscling and wholesale product yield traits. J. Anim. Sci. 76: 2255-2262.

ROTHSCHILD M.F., 1994. Genome mapping in livestock: A journey, not a destination. Proceedings of the Future Genetics for the Animal Industry, St. Louis, MO, May 4. ISU publication.

ROTHSCHILD M.F., 1999. Methods to determine genetic diversity. Discovery Conference Proceedings, Nashville, TN, November 2.

ROTHSCHILD M.F., JACOBSON C., VASKE D.A., TUGGLE C.K., WANG L., SHORT T., ERCHARDT G., SASAKI S., VINCENT A., MCLAREN D.G., SOUTHWOOD O., VAN DER STEEN H., MILEHAM A., PLASTOW G., 1996. The Estrogen receptor locus is associated with a major gene influencing litter size in pigs. Proc. National Acad. Sci. 93: 201-205.

ROTHSCHILD M.F., LIU H.C., TUGGLE C.K., YU T.P., WANG L., 1995. Analysis of pig chromosome 7 genetic markers for growth and carcass performance traits. J. Anim. Breed. Genet. 112: 341-348.

ROTHSCHILD M.F., PLASTOW G.S., 1999. Advances in pig genomics and industry applications. AgBiotechNet, 10: 1-8.

SASAKI S., CLUTTER A.C., POMP D., 1996. Assignment of the porcine obese (leptin) gene to chromosome 18 by linkage analysis of a new PCR-based polymorphism. Mamm. Genome 7: 471-472.

SHORT T.H., ROTHSCHILD M.F., SOUTHWOOD O.I., MCLAREN D.G., DEVRIES A., VAN DER STEEN H., ECKARDT G.R., TUGGLE C.K., HELM J., VASKE D.A., MILEHAM A.J., PLASTOW G.S, 1997a. Effect of the Estrogen receptor locus on reproduction and pro-

duction traits in four commercial pig lines. J. Anim. Sci. 75: 3138-3142.

SHORT T.H., SOUTHWOOD O.I., DEVRIES A., MCLAREN D.G., EVANS G.J., MILEHAM A.J., PLASTOW G.S., 1997b. Evidence of a new genetic marker for litter size in pigs. J. Anim. Sci. 75 (Suppl. 1): 29.

SIGNER E.N., ARMOUR J.A.L., JEFFREYS A.J., 1996. Detection of an MboI RFLP at the porcine clotting factor IX locus and verification of sex linkage. Anim. Genet. 27: 130.

SOUMILLION A., ERKENS J.H., LENSTRA J.A., RETTENBERGE, G., TE PAS, M.F.W., 1997. Genetic variation in the porcine myogenin gene locus. Mammalian Genome, 8: 564.

SUN H.S., WANG L., ROTHSCHILD M.F., TUGGLE C.K., 1998. Mapping of the natural-resistance associated macrophage protein (NRAMP1) gene in pigs. Anim. Genet. 29: 138-140.

TEPAS M.F.W., SOUMILLION A., RETTENBERGER G., VAN DEN BOSCH T.J., VENINGA G., MEUWISSEN T.H.E., 1996. Characterization of the porcine myogenin gene locus and association between polymorphisms and growth traits. Anim. Genet. 27 (Suppl. 2): 117.

VAN DER STEEN H.A.M., SOUTHWOOD O.I.S., DE VRIES A.G., SHORT T.H., MCLAREN D.G., WEI M., PLASTOW G.S., 1997. Evidence of a new genetic marker for litter size in meishan synthetic pigs. (ICAB Proceedings)

VAN OERS P.A.M., DE VRIES B.J., VAN DER POEL J.J., GROENEN M.A.M., 1996. Mapping QTL in pigs: the Dutch Meishan X White resource family. Anim. Genet 27 (Suppl. 2): 117.

VINCENT A.L., WANG L., ROTHSCHILD M.F., 1997. Linkage mapping of the leptin receptor gene to pig chromosome 6. J. Anim. Sci. 75: 2287.

VINCENT A.L., SHORT T.H., ECKARDT G.R., MCLAREN D.G., SOUTHWOOD O.I., PLASTOW G.S., TUGGLE C.K., ROTHSCHILD M.F., 1998. The prolactin gene receptor is associated with increased litter size in pigs. Proc 6World Congress Quant. Genetics of Livestock 27: 15-18.

WALLING G.A., VISSCHER P.M., ANDERSSON L., ROTHSCHILD M.F., WANG L., MOSER G., GROENEN M.A.M., BIDANEL J-P., CEPICA S., ARCHIBALD A.L., GELDERMANN H., DE KONING D.J., MILAN D., HALEY C.S., 2000. Combined analyses of data from QTL mapping studies: Chromosome 4 effects on porcine growth and fatness. Genet. (in press).

WANG L., YU T.P., TUGGLE C.K., ROTHSCHILD M.F., 1997a. Association of porcine chromosome 4 markers with growth and carcass traits. Plant & Animal Genome V Meetings Proceedings.

WANG L., YU T.P., CHUNG H.W., TUGGLE C.K., ROTHSCHILD M.F., 1998. A directed search for quantitative trait loci on chromosomes 4 and 7 in the pig. J. Anim. Sci. 76: 2560-2567.

WEBB A.J., 1998. Objectives and strategies in pig improvement: An applied perspective. J. Dairy Sci. 81 (suppl. 2) 36-46.

WILKIE P.J., PASZEK A.A., FLINKENGER G.H., ROHRER G.A., ALEXANDER L.J., BEATTIE C.W., SCHOOK L., 1996. Scan of eight porcine chromosomes for growth, carcass and reproductive traits reveals two likely quantitative trait loci. Anim. Genetics 27 (Suppl. 2): 117.

YU T.-P., TUGGLE C.K., SCHMITZ C.B., ROTHSCHILD, M.F., 1995. Association of PIT1 polymorphisms with growth and carcass traits in pigs. J. Anim. Sci. 73: 1282-1288.

YU T.P., WANG L., TUGGLE C.K., ROTHSCHILD M.F., 1998. Mapping genes for fatness and growth on pig chromosome 13: A search in the region close to the pig PIT1 gene. J. Anim. Breeding & Genet. 116: 269-280.

ITEA (2000), Vol. **96A N.° 3,** 362-369

EL ESQUEMA DE SELECCIÓN DE LA RAZA CAPRINA "CABRA DEL GUADARRAMA"

M. Serrano*, P. Díez de Tejada, D. Hernández*, J.J. Jurado***

* Departamento de Mejora Genética y Biotecnología, INIA, Ctra. de la Coruña km 7, 28040 Madrid, España

**Asociación de Criadores de raza de Guadarrama, Complejo Agropecuario de Colmenar Viejo, Ctra de Guadalix de la Sierra, km 2, Colmenar Viejo, España

RESUMEN

En este estudio se realiza una descripción general de la situación del ganado caprino de raza del Guadarrama y de las acciones que en el contexto de la mejora genética viene realizando el Departamento de Mejora Genética y Biotecnología del INIA de Madrid desde el año 1997 en colaboración con la Asociación de Criadores de raza de Guadarrama. Se describen los parámetros productivos de esta raza en cuanto a caracteres relacionados con la producción lechera. Se presentan los resultados obtenidos hasta la fecha en cuanto a modelización de los caracteres objeto de selección y valoración genética de los animales de esta raza.

Palabras clave: Cabras de Guadarrama, Producción lechera, Valoración genética.

SUMMARY

In this study a general description of Guadarrama goats breed and actions in animal breeding carried out by the Genetic Breeding and Biotechnology Department of INIA (Madrid, Spain) in collaboration with the Guadarrama Breeders Asociation since 1997, were described. Productive parameters related with milk traits were studied. Results related with trait modellization and genetic values prediction of animals of Guadarrama breed were presented.

Key words: Guadarrama goats, Milk traits, Genetic evaluation.

Introducción

El sector del ganado caprino no es tan importante en cuanto al número de cabezas, 2.725.000 en 1999 (Anuario Estadística Agroalimentaria. M.A.P.A.), como en cuanto al número de familias que subsisten de su explotación. Se trata en general del sector más marginal dentro de la producción animal, ya que suele desarrollarse la cría de la cabra en lugares donde otras especies no son rentables.

La raza caprina del Guadarrama también conocida por los nombres de "Guadarrameña", "Del Moncayo y Guadarrama" y "Cabra del Centro de España" está ampliamente implantada en la Comunidad de

Madrid, principalmente en el Sistema Central: Sierra de Guadarrama, Puerto de Somosierra y Sierra de Ayllón. Existen también efectivos de esta raza en la Sierra de Malagón que pertenece a la provincia de Ávila. Esta raza está considerada por la Comunidad Europea, M.A.P.A. y Comunidad de Madrid como *Raza en Peligro de Extinción* y recibe subvenciones de éstos organismos para su conservación y mejora.

Esta raza procede del tronco pirenaico y ha sido explotada tradicionalmente en su doble aptitud carne-leche, aunque con una clara orientación a la producción láctea, en un medio ecológico típico de las serranías del Macizo Central, con veranos cortos y frescos e inviernos largos y fríos, con abundancia de lluvias y nieve, en cuyo habitat se desenvuelve durante todo el año. La cabra del Guadarrama se adapta perfectamente a este medio desfavorable dada su elevada rusticidad, que le permite aprovechar los escasos recursos alimenticios existentes, moverse en terrenos difíciles y soportar las bajas temperaturas y lluvias. Por el contrario es sensible a las temperaturas elevadas cuando los animales se trasladan a zonas cálidas.

Morfológicamente se admiten cuatro capas en La cabra del Guadarrama: Vegata, Jardasca, Cárdena y Negra. Los animales se caracterizan por su gran rusticidad y su elevada capacidad lechera. Los sistemas de producción predominantes son el extensivo y el semi-extensivo. El sistema de pastoreo estante es el más frecuente, usándose cada vez menos el pastoreo transterminante. Al igual que otras razas caprinas de la Península, son poliéstricas permanentemente, si bien los partos, generalmente, tienen lugar en dos épocas del año: la más frecuente que coincide con los meses de octubre y noviembre, y otra, de menor incidencia, en los meses de febrero y marzo. Se produce un solo parto al año. Las hembras se cubren por primera vez a los seis-siete meses de edad. El índice de prolificidad es variable, si bien se puede estimar en 1,67.

La fuente de ingreso principal de las explotaciones caprinas de esta raza se basa fundamentalmente en la producción láctea. Una característica de esta agrupación es la ausencia de brucelosis, por lo que están permitidos los quesos realizados con leche cruda. La producción de cabritos es una actividad secundaria pero no por ello menos importante. Generalmente los cabritos se sacrifican a los 30-40 días de edad con 9 a 10 kg de peso vivo.

El censo de esta raza se puede estimar en 7.500 reproductoras inscritas en el Libro Genealógico pertenecientes a 33 explotaciones.

Desde el año 1997 el Departamento de Mejora Genética y Biotecnología del INIA viene colaborando con la Asociación de Criadores de Raza del Guadarrama, en el establecimiento de un esquema de selección para incrementar la producción lechera en los animales de esta raza. Hasta la fecha se han realizado dos Relaciones de Reproductores en 1998 y 1999, que, dada la ausencia de registros genealógicos, consisten en una corrección del valor fenotípico de producción de leche estandarizada a 210 días de lactación y 2,8% de proteína por una serie de efectos ambientales previamente analizados en modelos fijos.

En abril de 2000, se ha realizado la primera valoración genética intrarebaño de los animales de esta raza para varios caracteres relacionados con la producción lechera, ya que se cuenta con algunos registros genealógicos maternos. Los valores genéticos predichos para los animales, solo serán comparables en cada ganadería, debido a la desconexión existente entre éstas.

Actualmente se están llevando a cabo las primeras inseminaciones artificiales de las hembras de esta raza, con el fin de conectar ganaderías y probar machos por la producción de sus hijas en distintos rebaños. Esto permitirá en unos años, disponer de genealogías paternas y de una cierta conexión entre los rebaños adscritos al esquema de selección. Hasta la fecha se han realizado 300 inseminaciones con una fertilidad media del 50%.

La Comunidad de Madrid mediante un convenio entre el Instituto Tecnológico de Desarrollo Agrario (ITDA), el Instituto Madrileño de Investigaciones Agrarias (IMIA), la Asociación de Criadores de raza de Guadarrama y la Industria transformadora, está abordando el estudio y caracterización de los productos derivados de esta raza, queso y cabrito, con el fin de tipificarlos para poder obtener una Denominación de Origen que los ampare.

Material y métodos

Desde el año 1997 la asociación de criadores de raza caprina del Guadarrama, nos aporta los datos de control lechero de esta raza y las lactaciones tipificadas a 210 días de caracteres relacionados con la producción lechera. Hasta el año 2000 no se ha contado con resgistros genealógicos.

En 1998 se contaba con 22.655 controles lecheros pertenecientes a 4.270 cabras distribuídas en 26 rebaños y 1.785 lactaciones válidas. Este año se realizó una corrección fenotípica del carácter producción de leche estandarizada a 210 días de lactación y 2,8% de proteína en función de una serie de efectos ambientales previamente estudiados mediante una metodología de modelos fijos. El modelo de corrección de la producción de leche incluyó como efectos el rebaño-época de parto (ya que solo se disponía de datos desde agosto del 96 hasta julio del 97), donde los meses de parto se distribuyeron en tres clases (1 = septiembre a marzo; 2 = abril a mayo y 3 = agosto) de frecuencia homogénea; el intervalo entre el parto y el primer control (1 = < 44 días; 2 = 45-62 días; 3 = 63-89 días; 4 = > 89 días) y el número de parto (1 = 1.° parto; 2 = 2.° parto y 9 = desconocido). Las producciones corregidas sólo fueron comparables entre animales de un mismo rebaño, por la desconexión existente entre éstos.

En 1999 el control lechero de esta raza contaba con 47.632 controles. La base de datos con la que se realizó la corrección fenotípica para la producción de leche a 210 días y 2,8% de proteína constaba de 3.247 lactaciones estandarizadas válidas pertenecientes a 3.059 cabras distribuidas en 28 rebaños. El modelo de corrección fenótipica incluyó los efectos ambientales rebaño-año-época de parto (en este caso ya se disponía de registros de tres años, (1996, 1997 y 1998), con las estaciones distribuidas en 3 clases (enero-marzo, abril-septiembre y octubre-diciembre); el intervalo entre el parto y el primer control (4 niveles como en el caso anterior); el número de lactación (3 niveles igual que en el caso anterior) y el número de crías al parto (1 = una cría; 2 = 2 o más crías y 9 = desconocido). Al igual que en el caso anterior, las producciones corregidas sólo fueron comparables entre animales de un mismo rebaño.

En el año 2000 disponemos de una base de 84.461 controles lecheros. De estos se pudieron obtener 8.241 lactaciones válidas pertenecientes a 5.371 cabras distribuidas en 32 rebaños, con las que se ha realizado la primera valoración genética intrarebaño de los animales de esta raza, por disponer ya de

algunos registros genealógicos maternos. Previamente a la valoración genética se realizaron estudios de modelos fijos mediante el procedimiento GLM del SAS (SAS procedure, 1995), con el fin de analizar la significación de los factores ambientales ya definidos y la distribución de sus niveles.

Los caracteres analizados fueron kg de leche estandarizada a 210 días de lactación y 2,8% de proteína (KLEC), kg de grasa estandarizada a 210 días de lactación (KGRA), kg de proteína estandarizada a 210 días de lactación (KPRO), kg de lactosa estandarizados a 210 días de lactación (KLAC) y kg de extracto seco estandarizados a 210 días de lactación (KEXT).

La evaluación genética de los animales se realizó para los caracteres KLEC, KGRA y KPRO, con estimas de parámetros provenientes de referencias bibliográficas. Se utilizó la metodología BLUP con un modelo animal de medidas repetidas. El archivo de genealogía incluyó como únicas relaciones de parentesco conocidas 324 madres de cabras con dato productivo (263 de estas madres con dato propio para una o más lactaciones).

El modelo de valoración para los tres caracteres finalmente elegido fue aquel que incluyó como efectos fijos el grupo de comparación rebaño-año-época de parto (199 niveles: 32 rebaños, 4 años y 3 épocas de parto definidas como en el año 1999); intervalo entre el parto y el primer control (como en los casos anteriores); número de lactación (4 niveles al incluir tambien 3.ª lactación) y el número de crías al parto (3 niveles al incluir 3 o más crías como nivel independiente). Todos estos efectos se mostraron altamente significativos. Los efectos aleatorios del modelo fueron el genotipo del animal y el efecto ambiental permanente.

La ecuación del modelo animal utilizado en la valoración genética fue el siguiente:

$$y_{ijklmo} = \mu + rae_j + inte_k + lac_l + ncp_m + g_o + p_o + e_{ijklmo}$$

Donde:

y = observación *i* de producción de leche, grasa o proteína estandarizada a 210 días del *l* número de parto (lac) de la cabra *o* , perteneciente al *j* grupo de comparación (rae), con un intervalo parto primer control (inte) *k* y *m* número de crías al parto (ncp), 8241 registros.

rae = grupo de comparación rebaño-año-época de parto (199 niveles).

inte = intervalo parto-primer control (4 niveles).

lac = número de parto o lactación (4 niveles).

ncp = número de crías al parto (4 niveles).

g = genotipo del animal (5.431 niveles).

p = efecto ambiental permanente (5.371 niveles).

e = residuo.

La valoración genética de los caracteres se llevó a cabo mediante el programa estadístico BLUPAM (JURADO y col., 1991). Este programa obtiene las fiabilidades de las predicciones genéticas mediante el método aproximado desarrollado por Meyer (MEYER, 1989). La fiabilidad se expresa como porcentaje del cuadrado de la correlación entre el valor genético predicho y el esperado para cada animal valorado.

Los parámetros genéticos utilizados en la valoración de dichos caracteres fueron obtenidos de referencias bibliográficas. Las heredabilidades utilizadas en la valoración de KLEC, KGRA y KPRO fueron de 0,20,

0,32 y 0,31, respectivamente (BOICHARD y col., 1999). Las repetibilidades, dada la ausencia de referencias en caprino, se tomaron de estimas obtenidas en la especie ovina, y fueron de 0,40 para todos los caracteres.

Resultados y discusión

La media, desviación típica, mínimo y máximo de los caracteres analizados se presentan en el cuadro 1.

La producción de kilos de leche estandarizada a 210 días de lactación en la raza caprina Murciano-Granadina es, para animales con más de un parto, de 439 kg. según RABAL y col. (1998). Datos más recientes analizados en el año 2000 por Hernández (comunicación personal) arrojan una producción de 413,68 kg de leche a 210 días en animales de una o más lactaciones de esta misma raza, lo que supone 63 kg más de producción que la raza de Guadarrama (350 kg).

Estas diferencias en producción se explicarían por la distinta aptitud de los animales a la producción láctea. La raza Murciano-Granadina ha sido explotada tradicionalmente para la producción de leche, para lo cual han sido adaptados sus sistemas de producción. De este modo es frecuente en esta raza la lactancia artificial de los cabritos con el fin de ordeñar más tiempo a los animales. Sin embargo las cabras de raza de Guadarrama son aprovechadas en su doble aptitud carne-leche, de modo que la obtención de cabritos es una actividad importante en la economía de estas ganaderías. Otras razas caprinas españolas como la Malagueña (HERRERA y col. 1985) y Canaria (MÉNDEZ y RICO, 1992) presentan producciones de leche similares a las de la Murciano-Granadina y por tanto muy superiores a la de la raza del Guadarrama.

Estas razas también tienen un claro aprovechamiento lechero. La raza Verata (RODRÍGUEZ y col. 1988) presenta, sin embargo, producciones similares a la cabra de Guadarrama siendo éstas de 362 kg de leche en 210 días de lactación.

Cuadro 1

Media, desviación típica (DT), mínimo y máximo de los caracteres estandarizados (en kg) a 210 días de lactación, leche a 2,8% de proteína (KLEC), grasa (KGRA), proteína (KPRO), lactosa (KLAC) y extracto seco (KEXT)

Table 1

Mean, standar deviation (DT), minimun and maximun of standardised traits (in kg.) to 210 days in milk, milk yield adjusted to 2,8% of protein (KLEC), fat yield (KGRA), protein yield (KPRO), lactose yield (KLAC) and dry matter yield (KEXT)

Carácter	N	Media	DT	Mínimo	Máximo
KLEC	8.241	349.166	136.672	15.160	964.730
KGRA	8.241	12.923	5.358	0.770	40.330
KPRO	8.241	9.776	3.827	0.420	27.010
KLAC	8.241	13.153	5.307	0.600	37.810
KEXT	8.241	25.340	10.028	1.120	70.860

La media, desviación típica, mínimo y máximo de los porcentajes de grasa, proteína, lactosa y extracto seco estandarizados a 210 días de lactación para la raza caprina del Guadarrama se presentan en el cuadro 2.

Estos valores son inferiores a los encontrados en la raza Murciano-Granadina (HÉRNANDEZ, comunicación personal) que presenta porcentajes de 5,2, y 14,9 para grasa y extracto seco estandarizados a 210 días de lactación, respectivamente. Sin embargo los porcentajes de proteína son muy similares, 3,4 y 3,3 para las razas Murciano-Granadina y de Guadarrama, respectivamente, lo cual indica la aptitud quesera de la leche producida por ésta última.

Con respecto a las razas Canaria (MÉNDEZ y RICO, 1992) y Malagueña (HERRERA y col., 1985), los porcentajes de grasa de éstas son superiores a los encontrados en la cabra de Guadarrama, alcanzando valores de 5,4 y 5,1, respectivamente. Sin embargo la raza Verata (RODRÍGUEZ y col., 1988) presenta un porcentaje de grasa similar al estimado en la raza de Guadarrama, 4,3. El porcentaje de proteína solo es superior en el caso de la cabra Canaria en la que toma un valor de 4,7. Sin embargo tanto las razas Malagueña como Verata, presentan unos porcentajes proteicos muy similares al de la raza de Guadarrama, con valores de 3,3 y 3,4, respectivamente.

La evolución de la producción de los kilogramos de leche, grasa y proteína estandarizados a 210 días desde el año 1996 hasta 1999 para los animales controlados de raza del Guadarrama, se presenta en la figura 1.

Teniendo en cuenta que en los años 1996 y 1999 sólo se dispone de 968 y 1136 registros productivos, respectivamente, frente a los 2.885 de 1997 y 3.252 de 1998, años más representativos en cuanto a cantidad de información disponible, parece existir una ligera tendencia a la disminución de los kg de leche, grasa y proteína producidos por los animales de esta raza. Esta disminución podría deberse a factores ambientales.

En el cuadro 3 se presentan la desviación típica, mínimo y máximo de los valo-

Cuadro 2

Media, desviación típica (DT), mínimo y máximo de los caracteres estandarizados a 210 días de lactación, porcentaje de grasa (%GRA), porcentaje de proteína (%PRO), porcentaje de lactosa (%LAC) y porcentaje de extracto (%EXT)

Table 2

Mean, standard deviation (DT), minimun and maximun of standardised traits to 210 days in milk, fat content (%GRA), protein content (%PRO), lactose content (%LAC) and dry matter content (%EXT)

Carácter	N	Media	D.T.	Mínimo	Máximo
%GRA	8.241	4.430	0,862	0,55	8,91
%PRO	8.241	3.340	0,417	0,28	7,34
%LAC	8.241	4.464	0,433	0,43	6,56
%EXT	8.241	8.615	0,785	0,79	24,34

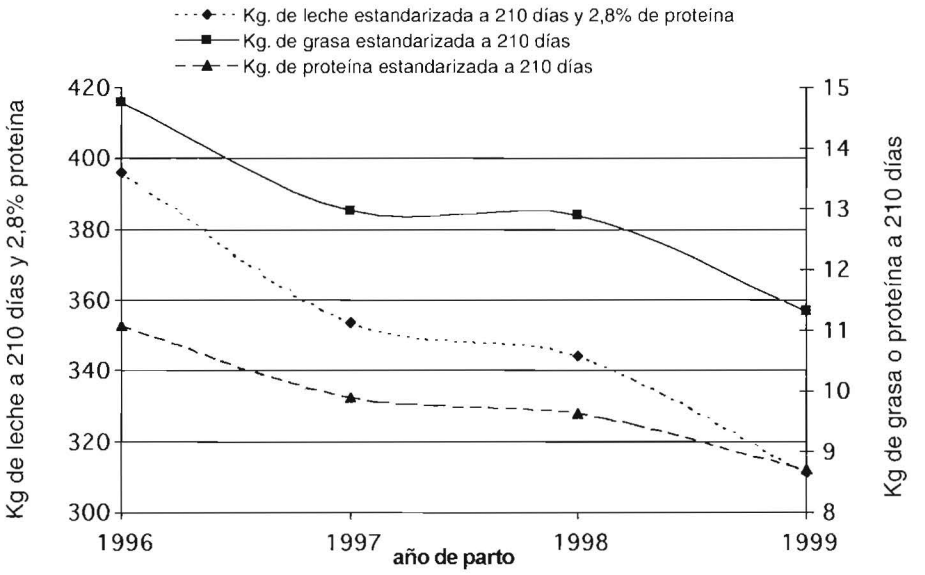


Figura 1. Evolución de los kg de leche, grasa y proteína estandarizados a 210 días de lactación, en el periodo 1996-1999

Figura 1. Evolution of milk, fat and protein yields (in Kg.) standardised to 210 days in milk, between years 1996 and 1999

res genéticos predichos y su fiabilidad para el conjunto de animales de la población.

La media del valor genético del 10% mejor de hembras para el carácter kg. de leche a 210 días y 2,8% de proteína es de +40.905 (540 observaciones). Esto supone casi dos desviaciones típicas (D.T.=20.954) con respecto a la media genética de la población (0 por definición) para dicho carácter. Así pues, aunque deberá ser confirmado en posteriores análisis de componentes de varianza, parece existir variabilidad genética suficiente para garantizar el progreso genético resultado de la selección de los animales para incrementar la producción de leche.

La fiabilidad de las predicciones genéticas presenta un amplio rango de variación, desde valores de 6,25 correspondientes a hembras sin dato propio pero que poseen una hija con registro para una lactación, hasta valores de 54,54 correspondientes a hembras con registro de tres lactaciones que tienen a su vez tres hijas con registro para una lactación. La fiabilidad media es de 32,65 que corresponde a hembras con registro conocido para una lactación que tienen madre conocida con registro para tres lactaciones y una hermana con registros de dos lactaciones. El valor más frecuente de fiabilidad (moda) es 25, que corresponde a hembras con un solo registro de lactación, sin relaciones de parentesco, madre o hijas, que aporten información.

Cuadro 3

Desviación típica (DT), mínimo y máximo de los valores genéticos y fiabilidad (FIAB) estimados para los caracteres estandarizados (en kg) a 210 días de lactación, leche a 2,8% de proteína (KLEC), grasa (KGRA) y proteína (KPRO). La fiabilidad es la estimada para el carácter KLEC

Table 2

Standard deviation (DT), minimun and maximun of genetic values and accuracy estimated for standardised traits (in kg.) to 210 days in milk, milk yield adjusted to 2,8% of protein (KLEC), fat yield (KGRA) and protein yield (KPRO). The accuracy is the estimated for KLEC

Carácter	N	D.T.	Mínimo	Máximo
KLEC	5.431	20,954	-60.830	97.045
FIAB	5.431	9,307	6.250	54.540
KGRA	5.431	1,287	-4.040	6.710
KPRO	5.431	0,906	-2.647	4.210

Conclusiones

Los datos analizados en la raza caprina del Guadarrama permiten detectar un potencial importante en esta raza en cuanto a producción lechera, y la existencia de un material genético interesante para el desarrollo de un esquema de selección para el incremento de la cantidad de leche producida por los animales de esta raza caprina.

Bibliografía

BOICHARD D., BOUILOC N., RICORDEAU G., PIACERE A., BARILLET F., 1999. Genetic parameters for first lactation dairy traits in Alpine and Saanen goat breeds. Genet. Sél. Evol. 21: 205-215.

HERRERA M., SUBIRES J., APARICIO J.B., SOLANO J., SÁNCHEZ M., PEÑA F., 1985. Capacidad productiva de la cabra Malagueña en sistemas extensivos. Simposio Internacional de Explotación Caprina en Zonas Áridas. Fuerteventura.

JURADO J.J., HERNÁNDEZ D., SERRANO M., 1991. II Catálogo de Software de Interés en Agricultura. Fundesco, IRYDA, MAPA. Programa 248 BLUPAM, pág. 142.

MÉNDEZ C., RICO A., 1992. Aproximación a la cabaña caprina canaria (II). Mundo Ganadero 1992-9, págs. 82-85.

MEYER K., 1989. Approximate accuracy of genetic evaluation under an animal model. Livest. Prod. Sci. 21: 87.

RABAL F., CORRALES J.C., SÁNCHEZ A., 1998. Resultados productivos de la cabra Murciano-Granadina. Mundo ganadero. Octubre, págs. 32-34.

RÓDRIGUEZ P., TOVAR J., DÍAZ A., 1988. Producción de leche de la cabra Verata. A.Y.M.A. 28: 51-55.

SAS Statistical Package, 1995. Sas Institute Inc. Cary, NC. USA.

ITEA (2000), Vol. **96A N.° 3,** 370-378

ESTIMACIÓN DEL PARENTESCO EN CERDOS IBÉRICOS UTILIZANDO MARCADORES MOLECULARES

M.A. Toro, C. Barragán, J. Rodrigáñez, C. Rodríguez, L. Silió

Departamento de Mejora Genética y Biotecnología, SGIT-INIA, Madrid, España

RESUMEN

Se han comparado distintos estimadores del parentesco obtenidos a partir de información molecular de 63 cerdos ibéricos procedentes de las estirpes Torbiscal y Guadyerbas para las que existen datos genealógicos desde 1945. Los marcadores utilizados fueron 32 microsatélites distribuidos entre los 18 autosomas. Los estimadores calculados fueron los siguientes: 1) L&H1 y L&H2 propuestos por LI y HORVITZ (1953); 2) L&R propuesto por LYNCH y RITLAND (1999); 3) Q&G propuesto por QUELLER y GOODNIGHT (1989); 4) RIT, propuesto por RITLAND (1996). Asimismo, se calculó el parentesco molecular y el genealógico, este último a partir del pedigree completo. Se detectaron fuertes discrepancias entre las estimaciones obtenidas a partir de la información molecular y las obtenidas en el análisis genealógico siendo los valores de las primeras mucho menores. También se evaluó la correlación entre los estimadores y los valores genealógicos. Esta correlación fue alta, entre 0,91 y 0,74 para los distintos estimadores, si se consideran los 63 animales que muestran un rango de parentescos muy extremo, pero disminuye hasta alcanzar valores de 0,60 y 0,30 para las poblaciones Guadyerbas y Torbiscal, respectivamente, cuando se analizan por separado. Por últimos se discuten las posibles causas de las discrepancias observadas.

Palabras clave: Microsatélites. Cerdo ibérico, Parentesco.

SUMMARY

Several estimators of coancestry from molecular information were compared. The data came from 63 iberian pigs of the strains Guadyerbas and Torbiscal with the complete genealogy known since 1945. The markers used were 32 microsatellites distributed among the 18 autosomes. The estimators calculated were: 1) L&H1 y L&H2 proposed by LI and HORVITZ (1953); 2) L&R proposed by LYNCH and RITLAND (1999); 3) Q&G proposed by QUELLER and GOODNIGHT (1989); 4) RIT, proposed by RITLAND (1996). Besides, the molecular and the genealogical coancestry was also calculated, the last one using the complete pedigree. Important discrepancies were found between the estimations obtained with molecular information and the coancestries calculated in the genalogical analysis, being the first much smaller. The correlation between the estimations and the genealogical values was also calculated. This correlation was high, between 0.91 and 0.74 for the different estimators when all the animals were considered and therefore a wide range of coancestries ocurred. However, the correlation decreases until values of 0.60 and 0.30 for the Guadyerbas and Torbiscal populations, respectively when they were analysed separately. Finally, the possible causes of the discrepancies are also discussed.

Key words: Microsatellites, Iberian pig, Coancestry.

Introducción

Los conceptos de consanguinidad de un individuo y de parentesco entre dos individuos juegan un papel de primordial importancia en las aplicaciones prácticas de la Genética cuantitativa. Así, dichos coeficientes se requieren tanto para estimar los parámetros genéticos como para realizar evaluaciones del mérito genético de aplicación en la Mejora. En Sociobiología son también importantes para interpretar la evolución de los comportamientos sociales. Y por último, en la Genética de la conservación constituyen la herramienta fundamental para la toma de decisiones tanto respecto a los individuos que han de contribuir genes a la siguiente generación como respecto a la forma en que deben ser apareados (CABALLERO y TORO, 2000).

El espectacular desarrollo de los marcadores moleculares ocurrido en los últimos años ha motivado un creciente interés en utilizarlos para determinar el parentesco (AVISE, 1994). Se han publicado ya varias aplicaciones en la estimación de parámetros genéticos en poblaciones naturales de plantas (RITLAND, 1996) y en la estimación del parentesco entre colonias en los insectos sociales (QUELLER y GOODNIGHT, 1989).

Se han propuesto numerosos estimadores del parentesco a partir de la información proporcionada por los marcadores la mayoría de los cuales se describen en la reciente revisión de LYNCH y RITLAND (1999). Su calidad como estimadores del verdadero parentesco se ha establecido fundamentalmente mediante simulación concluyéndose que proporcionan estimas razonables. Sin embargo estas simulaciones suponen una estructura de parentesco muy sencilla, básicamente padres e hijos, hermanos y medios hermanos.

En el presente trabajo se comparan distintos estimadores del parentesco en una población de cerdos ibéricos de genealogía compleja y totalmente conocida, de forma que pueden compararse los resultados de los distintos estimadores entre sí y con los correspondientes valores obtenidos del pedigree.

Materiales y métodos

Poblaciones

Las dos poblaciones consideradas en este estudio Guadyerbas y Torbiscal forman parte de la piara experimental de cerdos ibéricos establecida desde 1944-1945 en la finca Dehesón del Encinar (Oropesa, Toledo). La formación de esta piara se realizó a partir de animales procedentes de cuatro estirpes representativas de los principales tipos de cerdo ibérico existentes en la época: dos portuguesas coloradas: *Ervideira* y *Caldeira* y dos extremeñas negro-lampiñas: *Campanario* y *Puebla* (ODRIOZOLA, 1976). *Guadyerbas* es la actual denominación de la estirpe *Puebla* mientras que *Torbiscal* es una línea compuesta que procede de la fusión de las cuatro estirpes fundadoras realizada lentamente entre 1961 y 1963 (RODRIGÁÑEZ y col. 2000). Por lo tanto, las dos poblaciones comparten genealogía hasta 1963. Desde entonces han permanecido aisladas genéticamente, conservándose en la actualidad con censos aproximados de 7 machos y 28 hembras (*Guadyerbas*) y 15 machos y 60 hembras (*Torbiscal*). Se dispone de la genealogía completa de todos los animales desde 1945, habiendo transcurrido unas 20 generaciones hasta los actuales animales genotipados (TORO y col., 2000).

Microsatélites

Se utilizaron 32 microsatélites distribuidos entre los 18 autosomas de la forma siguiente: uno en los cromosomas 1, 2, 3, 5, 7, 9, 10, 11, 13, 16, 17 y 18, dos en los cromosomas 8, 14 y 15, tres en los cromosomas 6 y 12 y ocho en el cromosoma 4. Los microsatélites, amplificados mediante PCR, se analizaron por electroforesis capilar en un ABI 310. El rango en el número de alelos/marcador fue 2-7 con un número medio de 4,2 y el rango de heterocigosidades 0,18-0,79 siendo la heterocigosidad media 0,48.

Estimación del parentesco: análisis genealógico

El parentesco de todas las posibles parejas de individuos se calculó a partir del pedigree utilizando el método tabular. Con objeto de homogeneizar los resultados en lugar de calcular también la consanguinidad de cada individuo se calculó el coeficiente de autoparentesco de un individuo consigo mismo. La relación entre el coeficiente de autoparentesco y el de consanguinidad es simplemente $f_{XX} = 0,50\,(1 + F_X)$.

Estimación del parentesco con marcadores: un sólo locus

Se denomina parentesco molecular (MOL) al obtenido aplicando la definición de parentesco de MALÉCOT (1948) al gen marcador. Según esto el parentesco molecular entre dos individuos X e Y es la probabilidad de que dos alelos tomados al azar, uno de cada individuo sean idénticos en estado. Esta probabilidad puede también calcularse para cada uno de los alelos involucrados y el parentesco para ese locus será la suma de todos ellos. Estos valores calculados para las distintas parejas posibles de individuos en un locus multialélico se dan en el cuadro 1.

La relación entre el parentesco molecular y el de identidad por descendencia se entiende fácilmente si se considera que

$$E(f_{Mi}) = p_i^2 + p_i(1 - p_i)f$$

siendo f el coeficiente de parentesco de identidad por descendencia de Malécot y p_i la frecuencia del alelo *i*.

Según esto, una estima del parentesco contribuido por el alelo i será

$$f = \frac{f_{Mi} - p_i^2}{p_i(1 - p_i)}$$

Cuadro 1
Parentesco molecular para distintas parejas de invidulos
Table 1
Pairwise molecular coancestry

Pareja	f_{Mi}	f_{Mj}	f_{Mk}	f_M
A_iA_i, A_iA_i	1,00	0,00	0,00	1,00
A_iA_i, A_iA_j	0,50	0,00	0,00	0,50
A_iA_i, A_jA_k	0,00	0,00	0,00	0,00
A_iA_j, A_iA_j	0,25	0,25	0,00	0,50
A_iA_j, A_iA_k	0,25	0,00	0,00	0,25

La información proporcionada por n alelos puede combinarse de varias formas dando lugar a distintos estimadores. Si se elige como factor de ponderación $w_i = p_i(1 - p_i)$ como sugieren LI y HORVITZ (1953) se obtiene

$$f = \frac{\sum_i^n f_{Mi} - \sum_i^n p_i^2}{\sum_i^n p_i(1 - p_i)} = \frac{f_M - \sum_i^n p_i^2}{1 - \sum_i^n p_i^2} \qquad [1]$$

Si el factor de ponderación fuera $w_i = (1 - p_i)/(n - 1)$ como sugiere RITLAND (1996) suponiendo que las estimas son independientes, se obtiene

$$f = \sum_i^n \frac{f_{Mi} - p_i^2}{(n-1)p_i} = \frac{1}{(n-1)} \sum_i^n \frac{f_{Mi} - p_i^2}{p_i} \qquad [2]$$

Recientemente, LYNCH y RITLAND (1999) han propuesto el siguiente estimador de "regresión" (que llamaremos L&R) entre un individuo de referencia X (con alelos a y b) y un individuo probando Y (con alelos c y d).

$$f_{XY} = \frac{p_a(S_{bc} + S_{bd}) + p_b(S_{ac} + S_{ad}) - 4p_a p_b}{2(1 + S_{ab})(p_a + p_b) - 8p_a p_b}$$

siendo S_{uv} una variable auxiliar que toma el valor 1 si $u = v$ y 0 si $u \neq v$. Las dos estimas obtenidas para cada pareja f_{XY} y f_{YX} se promedian aritméticamente para obtener la estima final f.

Otro estimador relacionado con el anterior es el de QUELLER y GOODNIGHT (1989) que tiene la siguiente expresión:

$$f_{XY} = \frac{0,5(S_{ac} + S_{ad} + S_{bc} + S_{bd}) - p_a - p_b}{1 + S_{ab} - p_a - p_b},$$

y presenta la limitación de no estar definido para loci bialélicos.

Estimación del parentesco con marcadores: varios loci

A partir del estimador de LI y HORVITZ (1953) calculado para cada locus puede promediarse sin más para todos los N loci

$$f = \frac{\sum_l^N f_l}{N}$$

estimador que denominaremos L&H1 o bien promediarse primero el parentesco molecular de todos los loci y posteriormente aplicar (1) obteniéndose

$$f = \frac{\sum_l^N f_{Ml} - \sum_l^N \sum_{li}^n p_{li}^2}{1 - \sum_l^N \sum_{li}^n p_{li}^2}$$

estimador que denominaremos L&H2 y que, como indica RITLAND (1996) asume implícitamente que la correlación entre las estimas de los distintos loci es la unidad. Si por otra parte, se asumiera que dicha correlación es cero, la ponderación para cada locus sería $1/(n_l - 1)$ que aplicado sobre la ecuación (2) daría lugar al estimador que denominamos RIT:

$$f = \frac{\sum_l^N \sum_{li}^n \frac{f_{li} - p_{li}^2}{p_{li}}}{\sum_l^N (n_l - 1)}$$

En el caso del estimador de LYNCH y RITLAND (1999), que denominaremos L&R el parentesco multilocus se obtiene ponderando por el inverso de las varianzas unilocus:

$$f = \frac{1}{\sum_l^N w_l} \sum_l^N w_l f_l$$

$$\text{con } w_l = \frac{(1 + S_{ab})(p_a + p_b) - 4p_a p_b}{8p_a p_b}$$

Por último el estimador multilocus de QUELLER y GOODNIGHT (1989) es el promedio aritmético de las estimas unilocus y lo designaremos como Q&G.

Resultados

En el cuadro 2 se presentan una descripción de los coeficientes estudiados. Se ha distinguido entre los coeficientes de parentesco dentro de cada una de las dos poblacio-

Cuadro 2
Estadísticos de los estimadores del coeficiente de parentesco (f_{xy}) y autoparentesco (f_{xx}) calculados para individuos pertenenecientes a las líneas *Torbiscal* y *Guadyerbas*
Table 2
Statistics of the estimators of the coancestry coefficient (f_{xy}) and self-coancestry (f_{xx}) calculated for the individuals belonging to the Torbiscal *y* Guadyerbas *strains*

Estimadores Parentesco (f)	Guadyerbas		Torbiscal		Guad. vs Torb.
	fxy	*fxx*	*fxy*	*fxx*	*fxy*
(n)	(496)	(32)	(465)	(31)	(992)
Genealogía (PED)					
Media	0,390	0,677	0,163	0,574	0,047
Desviacion típica	0,032	0,061	0,030	0,010	0,0002
Min./ máx.	0,343/ 0,548	0,666/ 0,686	0,138/ 0,363	0,566/ 0,605	0,047/ 0,048
Molecular (MOL)					
Media	0,577	0,772	0,449	0,719	0,349
Desviacion típica	0,044	0,043	0,047	0,031	0,043
Min./ máx.	0,469/ 0,727	0,687/ 0,859	0.320- 0,711	0,641/ 0,766	0,242/ 0,492
Li y Horvitz (L&H1)					
Media	0,223	0,582	0,058	0,517	-0,156
Desviacion típica	0,091	0,090	0,088	0,055	0,076
Min./máx.	-0,061/ 0,516	0,416/ 0,764	-0,213/ 0,536	0,373/ 0,584	-0,458/ 0,158
Li y Horvitz (L&H2)					
Media	0,250	0,596	0,023	0,502	-0,153
Desviacion típica	0,078	0,076	0,084	0,056	0,076
Min./ máx.	0,058/ 0,515	0,446/ 0,751	-0,205/ 0,488	0,363/ 0,585	-0,343/ 0,100
Lynch y Ritland (L&R)					
Media	0,129	0,500	0,070	0,500	-0,113
Desviacion típica	0,072	0,000	0,057	0,000	0,036
Min./ máx.	-0,051/ 0,355	0,500/ 0,500	-0,048/ 0,426	0,500/ 0,500	-0,223/ -0,007
Queller y Goodnigth (Q&G)					
Media	0,222	0,500	0,011	0,500	-0,107
Desviacion típica	0,079	0,000	0,114	0,000	0,032
Min./ máx.	-0,015/ 0,418	0,500/ 0,500	-0,449/ 0,475	0,500/ 0,500	-0,299/ -0,006
Ritland (RIT)					
Media	0,094	0,396	0,091	0,683	-0,107
Desviacion típica	0,051	0,122	0,074	0,167	0,032
Min./ máx.	-0,036/ 0,289	0,252/	-0,068/ 0,519	0,394-1,107	-0,193/ 0,006

nes consideradas y el referido a individuos de poblaciones diferentes. Los coeficientes de autoparentesco genealógico medio son 0,677 y 0,574 que corresponden a coeficientes de consanguinidad promedio de 0,354 y 0,148 para *Guadyerbas* y *Torbiscal*, respectivamente. Los valores correspondientes de la homocigosidad (consanguinidad) molecular fueron 0,544 y 0,438. Los estimadores estudiados infraestiman en general el autoparentesco. Por otra parte tanto el estimador L&R como el Q&R asignan siempre un valor de 0,50 al autoparentesco. Los estimadores considerados también proporcionan un valor de

Cuadro 3

Correlación entre estimadores del coeficiente de parentesco (bajo la diagonal) y autoparentesco (sobre la diagonal) calculados para 63 individuos de las líneas *Guadyerbas* y *Torbiscal*

Table 3

Correlation among the estimators of coancestry coeffcint (below the diagonal) and self-coancestry (above the diagonal) calculated for 63 individuals of the Guadyerbas *y* Torbiscal *strains*

	PED	MOL	L&H1	L&H2	L&R	Q&G	RIT
Genealogía (PED)		0,581	0,394	0,581	0,000	0,000	-0,690
Molecular (MOL)	0,909		0,919	1,000	0,000	0,000	-0,249
Li and Horvitz (L&H1)	0,840	0,971		0,919	0,000	0,000	-0,125
Li and Horvitz (L&H2)	0,909	1,000	0,971		0,000	0,000	-0,249
Lynch y Ritland (L&R)	0,845	0,907	0,888	0,907		0,000	0,000
Queller y Goodnight (Q&G)	0,744	0,797	0,797	0,797	0,723		0,000
Ritland (RIT)	0,762	0,840	0,854	0,854	0,966	0,692	

Cuadro 4

Correlación entre estimadores del coeficiente de parentesco (bajo la diagonal) y autoparentesco (sobre la diagonal) calculados para 32 individuos de la línea *Guadyerbas*

Table 4

Correlation among the estimators of coancestry coefficient (below the diagonal) and self-coancestry (above the diagonal) calculated for 32 individuals of the Guadyerbas *strain*

	PED	MOL	L&H1	L&H2	L&R	Q&G	RIT
Genealogía (PED)		-0,216	-0,236	-0,216	0,000	0,000	-0,442
Molecular (MOL)	0,534		0,912	1,000	0,000	0,000	0,303
Li and Horvitz (L&H1)	0,431	0,899		0,912	0,000	0,000	0,247
Li and Horvitz (L&H2)	0,534	1,000	0,899		0,000	0,000	0,303
Lynch y Ritland (L&R)	0,625	0,735	0,600	0,735		0,000	0,000
Queller y Goodnight (Q&G)	0,438	0,593	0,523	0,593	0,649		0,000
Ritland (RIT)	0,646	0,739	0,593	0,739	0,946	0,635	

los coeficientes de parentesco muy inferior al valor obtenido en el análisis genealógico. Especialmente notable es el caso del parentesco entre animales de poblaciones diferentes para los que el promedio de las estimas es negativo e incluso para alguna de ellas todos los valores son negativos (L&R y Q&G).

En los cuadros 3, 4, 5 y 6 se presentan las correlaciones entre los valores de las distintas estimaciones tanto para el conjunto de todos los individuos como para cada población por separado. Para el conjunto de individuos la correlación más alta con el parentesco genealógico la presenta el parentesco molecular y el L&H2 del que difiere sólo en un cambio de escala siendo el valor de esta correlación próximo a 0,91. Con valores próximos a 0,85 se encuentran los estimadores L&H1 y L&R, mientras que los estimadores Q&G y RIT presentan los valores más bajos aunque todavía considerables, en torno a 0,75. También el autoparentesco genealógico presenta una correlación máxima con el molecular aunque el valor es más bajo (0,58). Los estimadores L&R y Q&G no proporcionan una estima y la que proporciona RIT esta correlacionada negativamente (-0,69).

Cuando se consideran las dos poblaciones por separado (cuadros 4 y 5) el orden de los estimadores respecto a su correlación con el parentesco genealógico es RIT,L&R, MOL = L&H2, Q&G yL&H1 en la población Guayerbas y L&R, RIT, MOL = L&H2, Q&G y L&H1 en Torbiscal. Sin embargo, llama la atención que incluso para los mejores estimadores los valores de la correlación son bastante más bajos, del orden de 0,60 y 0,30 en las poblaciones Guadyerbas y Torbiscal. Para el parentesco de animales de poblaciones distintas (cuadro 6) las correlaciones son próximas a cero y la mayoría negativas (de -0,045 a 0,038). Más paradójico todavía, el autoparentesco genealógico está negativamente correlacionado con todos los estimadores (~ -0,2) mientras que lo esta positivamente en Torbiscal (~ 0,3).

Cuadro 5
Correlación entre estimadores del coeficiente de parentesco (bajo la diagonal) y autoparentesco (sobre la diagonal) calculados para 31 individuos de la línea *Torbiscal*
Table 5
Correlation among the estiamators of coancestry coefficient (below the diagonal) and self-coancestry (above the diagonal) calculated for 31 individuals of the Torbiscal *strain*

	PED	MOL	L&H1	L&H2	L&R	Q&G	RIT
Genealogía (PED)		0,323	0,271	0,323	0,000	0,000	0,296
Molecular (MOL)	0,283		0,950	1,000	0,000	0,000	0,278
Li y Horvitz (L&H1)	0,265	0,972		0,950	0,000	0,000	0,273
Li y Horvitz (L&H2)	0,283	1,000	0,972		0,000	0,000	0,278
Lynch y Ritland (L&R)	0,329	0,769	0,704	0,770		0,000	0,000
Queller y Goodnight (Q&G)	0,268	0,564	0,563	0,564	0,568		0,000
Ritland (RIT)	0,301	0,690	0,622	0,690	0,966	0,532	

Cuadro 6
Correlación entre estimadores del coeficiente de parentesco calculados para 32 x 31 parejas de individuos de las líneas *Guadyerbas* y *Torbiscal*
Table 6
Correlation among the estimators of coancestry coefficient calculated for the 32 x 31 pairs of individuals of the Guadyerbas *y* Torbiscal *strains*

	PED	MOL	L&H1	L&H2	L&R	Q&G	RIT
Genealogía (PED)							
Molecular (MOL)	-0,035						
Li y Horvitz (L&H1)	-0,045	0,925					
Li y Horvitz (L&H2)	-0,035	1,000	0,925				
Lynch y Ritland (L&R)	-0,040	0,691	0,572	0,691			
Queller y Goodnight (Q&G)	0,038	0,413	0,371	0,413	0,200		
Ritland (RIT)	-0,042	0,878	0,744	0,878	0,827	0,396	

Discusión

El coeficiente de parentesco genealógico es un coeficiente que mide la probabilidad de identidad por descendencia relativo a una población de referencia (población base o fundadora) en la que se supone que todos los alelos de la misma son distintos. Por ello su valor es siempre mayor que cero. El coeficiente de parentesco molecular mide esencialmente desviaciones de las frecuencias genotípicas respecto al equilibrio de Hardy-Weinberg. La población de referencia, por tanto, sería una población hipotética con las mismas frecuencias génicas, pero en equilibrio de H-W.

La conversión de un coeficiente en otro, esto es la posibilidad de estimar el coefciente genealógico a partir de datos moleculares, requiere asumir que en la población fundadora considerada en el análisis genealógico los marcadores estaban tanto en equilibrio de H-W como en equilibrio de ligamiento y además con las mismas frecuencias que las que hemos observado en la población actual, esto es que no ha habido cambio en dichas frecuencias.

En una situación como la analizada en el presente trabajo no parece que pueda asumirse que las frecuencias actuales de alelos de los marcadores sean las mismas que las que había en la población base, debido al proceso de separación de las dos estirpes y a la deriva genética acumulada durante más de 40 años. Esta sería la explicación más plausible de las discrepancias entre las estimaciones obtenidas a partir de la información molecular y las obtenidas en el análisis genealógico. Las primeras son mucho menores y en muchos casos hay valores negativos, que indicarían una disminución de heterocigotos respecto a los que cabría esperar en el equilibrio. En el caso más extremo para el parentesco entre animales de distintas estirpes todos los valores de los estimadores L&R y Q&G son negativos.

En genética de la conservación el interés de utilizar información molecular está orientado a la elección de los animales menos emparentados para fundar un programa de

conservación o para tomar decisiones sobre la forma de aparearlos. En este contexto, es importante saber en que medida los estimadores basados en información molecular sirven como predictores de los parentescos genealógicos, esto es la magnitud de la correlación entre ambos. Los resultados de nuestro estudio indican que, para los 32 microsatélites utilizados, esta correlación es alta, entre 0,91 y 0,74 para los distintos estimadores si se consideran los 63 animales que muestran un rango de parentescos muy extremos (de 0,047 a 0,548). Sin embargo, si el rango de variación es menos extremo como en la población *Guadyerbas* (de 0, 343 a 0,548) o en la *Torbiscal* (de 0,138 a 0,363) el valor de la correlación es del orden de 0,60 y 0,30 respectivamente, llegando a ser negativa o próxima a cero cuando se consideran animales de poblaciones distintas. Lamentablemente la ordenación de los estimadores no fue siempre la mismo, ya que MOL fue el mejor en el conjunto de las dos poblaciones y RIT y L&R en las poblaciones *Guadyerbas* y *Torbiscal*, respectivamente.

A lo largo del trabajo hemos supuesto, al menos implícitamente, que el parentesco genealógico es el que mejor resume la verdadera variabilidad genética de todo el genoma y que el interés de los marcadores reside en proporcionar una estima, más o menos adecuada, del mismo. Obviamente conforme el abaratamiento y la robotización en la obtención de la información molecular facilite la obtención de marcadores, el parentesco molecular será en sí misma la medida más relevante.

Agradecimientos

Financiado por el INIA en el marco del proyecto SC98-083.

Bibliografía

CABALLERO A., TORO M.A., 2000. Interrelations between effective population size and other pedigree tools for the manegement of conserved populations. Genet. Res. 75: 331-343.

LI C.C., HORVITZ, D.G., 1953. Some methods of estimating the inbreeding coefficinent. Am. J. Hum. Genet. 5, 107-117.

LYNCH M., RITLAND K., 1999. Estimation of the pairwise relatedness with molecular markers. Genetics 152, 1753-1766.

MALECOT G., 1948. Les mathématiques de l'hérédité. Masson et Cie, Paris.

ODRIOZOLA M., 1976. Investigación sobre los datos acumulados en dos piaras experimentales. IRYDA. Madrid.

QUELLER D.C., GOODNIGHT K.F., 1989. Estimating relatedness using molecular markers. Evolution 43, 258-274.

RITLAND K., 1996. Estimators for pairwise relatedness and inbreeding coefficients. Genet. Res. 67, 175-186.

RODRIGÁÑEZ J., TORO M.A., RODRÍGUEZ C., SILIÓ L., 2000. Alleles survival from portuguese and spanish strains in a population of Iberian pig. Cahiers Options Mediteranéennés 41: 57-61.

TORO M.A., RODRIGÁÑEZ J., RODRÍGUEZ C., SILIÓ L., 2000. Genealogical analysis of a closed herd of black hairless Iberian pigs. Conservation Biology (en prensa).

ITEA (2000), Vol. **96A N.° 3**, 379-382

ANÁLISIS BAYESIANO MULTIVARIANTE DE COMPONENTES DE VARIANZA DE CARACTERES DE PRODUCCIÓN EN UNA POBLACIÓN PIETRAIN USANDO MARCADORES MOLECULARES

L. Varona*, G. Davalos, M. Pérez-Enciso*, J.M. Folch**, N. Jiménez**, A. Sánchez**, J.L. Noguera***

* Centro UdL-IRTA, Área de Producción Animal, 25198 Lleida, España

** Unidad de Genética y Mejora Animal, Universidad Autónoma de Barcelona, España

RESUMEN

Se ha llevado a cabo un análisis bayesiano multivariante para la estimación de parámetros genéticos de los caracteres peso y espesor de tocino dorsal a los 175 días en una población Pietrain. Se han utilizado los datos de 420 animales, que fueron genotipados para 9 marcadores moleculares. El pedigrí se completó con 5 machos y 58 hembras, también genotipados para los mismos marcadores. El modelo de análisis incluye: efectos sistemáticos (sexo y lote); un efecto genético poligénico; y cuatro efectos genéticos asociados a segmentos de 30 cM en los cromosomas 1, 2, 3 y 6. En el peso a los 175 días, las heredabilidades de origen poligénico (0,18) y asociadas a los cromosomas 4 (0,11) y 6 (0,15) fueron las más elevadas. Para espesor de tocino dorsal, son relevantes las heredabilidades poligénica (0,19), del segmento del cromosoma 2 (0,15) y del segmento del 6 (0.21). Las correlaciones genéticas entre los efectos atribuibles a los cromosomas 1, 3 y 6 fueron altas (> 0,65). Por el contrario, las correlaciones genéticas para el efecto poligénico y el cromosoma 2 fueron cercanas al cero.

Introducción

La creciente aparición de marcadores moleculares ha permitido el desarrollo de metodologías para la localización de genes que afectan a la expresión cuantitativa de ciertos caracteres de interés (QTLs), tanto en diseños de cruzamiento entre líneas (HALEY *et al.*, 1994), como en población abierta (GRIGNOLA *et al.*, 1996; VARONA *et al.*, 2000). Un enfoque alternativo consiste en identificar regiones del genoma que expliquen parte de la variación genética, tal y como ha sido sugerido por VISSCHER y HALEY (1996), LUI y DEKKERS (1998), VARONA y PÉREZ ENCISO (1998) y PÉREZ ENCISO y VARONA (2000). Este enfoque no necesita ninguna asunción acerca de las frecuencias génicas de los QTLs, ni siquiera intenta aportar información acerca del número y efecto de los genes involucrados.

El objetivo de este artículo es presentar una generalización del procedimiento presentado por VARONA y PÉREZ ENCISO (1998) a un análisis multicarácter. Para ello se han utilizado los datos de una población Pietrain,

controlada en el marco del proyecto UE PL 96-2243, uno de cuyos objetivos es evaluar si variaciones en regiones genómicas control localizadas en poblaciones experimentales también se encuentran en poblaciones comerciales

Material

Se utilizaron datos de peso (P) y espesor de tocino dorsal (ETD), registrados a los 175 días de edad, de 420 individuos distribuidos en 10 lotes de engorde, y procedentes de 5 machos y 58 hembras. Cada uno de los 483 individuos (parentales incluidos) fue genotipado para 9 marcadores genéticos: CGA y SW1430 en el cromosoma 1; S0141 y SW2623 en el cromosoma 2; SW732, S0206 y SW2618 en el cromosoma 3; S0003 y SW316 en el cromosoma 6.

Métodos

Los modelos de análisis de los datos de P (yw) y ETD (yb) fueron los siguientes:

$$y_w = X\beta_w + Zu_{wp} + Zu_{w1} + Zu_{w2} + Zu_{w3} + Zu_{w6} + e_w$$
$$y_b = X\beta_b + Zu_{bp} + Zu_{b1} + Zu_{b2} + Zu_{b3} + Zu_{b6} + e_b$$

donde y_w es el vector de datos de peso a los 175 días, y_b es el vector de datos de espesor de tocino dorsal, bw y bb son los vectores de efectos sistemáticos (sexo y lote), u_{wp} y u_{bp} son los efectos genéticos poligénicos, y u_{wi} y u_{bi} son los efectos genéticos asociados a un segmento de 30 cM en la región de los marcadores moleculares utilizados en el cromosoma i. Los vectores de residuos son e_w y e_b. Por último, X y Z son las matrices de incidencia que relacionan los registros fenotípicos con los variables del modelo.

Por la tanto la verosimilitud de los datos, dados los parámetros del modelo, se consideró una distribución normal multivariante:

$$f(y_w, y_b|\beta_w, \beta_b, u_{wp}, u_{bp}, u_{w1}, u_{b1}, u_{w2}, u_{b2},$$
$$N\left(\begin{matrix} X\beta_w + Zu_{wp} + Zu_{w1} + Zu_{w2} + Zu_{w3} + \\ X\beta_b + Zu_{bp} + Zu_{b1} + Zu_{b2} + Zu_{b3} + \end{matrix}\right.$$
$$\left.\begin{matrix} ,u_{w3}, u_{b3}, u_{w6}, u_{b6}) \propto \\ + Zu_{w6} \\ + Zu_{b6} \end{matrix}, R \otimes I\right)$$

donde R es la matriz de varianzas y covarianzas residuales.

Las distribuciones a priori de los parámetros fueron:

$$f(\beta_w, \beta_b) \propto k_1$$

$$f(u_{wp}, u_{bp}) = N(0, A_p \otimes G_p)$$

$$f(u_{w1}, u_{b1}) = N(0, A_1 \otimes G_1)$$

$$f(u_{w2}, u_{b2}) = N(0, A_2 \otimes G_2)$$

$$f(u_{w3}, u_{b3}) = N(0, A_3 \otimes G_3)$$

$$f(u_{w6}, u_{b6}) = N(0, A_6 \otimes G_6)$$

donde A_p es la matriz de relaciones genéticas para el efecto poligénico, y A_1, A_2, A_3 y A_6 son las matrices de relaciones genéticas para los efectos asociados a los segmentos de los cromosomas 1, 2, 3 y 6, respectivamente. Gp es la matriz de varianzas y covarianzas genéticas entre efectos poligénicos y G_p, G_1, G_2, G_3 y G_6 son las matrices de varianzas y covarianzas entre efectos asociados a los segmentos cromosómicos. Las distribuciones *a priori* de estas matrices de varianzas y covarianzas (G_p, G_1, G_2, G_3 y G_6) se asumieron planas.

Cuadro 1
Medias posteriores de las heredabilidades y correlaciones

		Cromosoma				
	Polig.	I	II	III	VI	Residual
h^2 P	0,18	0,08	0,09	0,11	0,15	-
h^2 ETD	0,19	0,07	0,15	0,10	0,21	-
r^2	0,06	0,79	-0,18	0,68	0,95	0,43

Las matrices de relaciones genéticas para el componente poligénico (Ap) se han calculado utilizando exclusivamente la genealogía, y las matrices de relaciones entre los segmentos cromosómicos (A_1, A_2, A_3 y A_6) se han calculado mediante el algoritmo de cadenas de Markov de Monte Carlo descrito por PÉREZ ENCISO *et al.* (2000).

A partir de la distribución conjunta que se obtiene de la verosimilitud y distribuciones a priori previamente descritas, se ha implementado un algoritmo de muestreo de Gibbs (GEMAN y GEMAN, 1984; WANG *et al.*, 1994), que involucra el muestreo de distribuciones condicionales univariantes normales para los parámetros de localización y de distribuciones *Wishart* invertidas para los parámetros de dispersión. Se realizó un análisis de cadena única que consistió en 250.000 iteraciones, de las cuales se descartaron las 50.000 primeras.

Resultados

En la cuadro 1 se presentan los resultados de las medias posteriores para las heredabilidades y las correlaciones genéticas y residuales. Para el carácter P, el cromosoma 6 representa el 15% de la variación genética y el componente poligénico el 18%. Por el contrario, para ETD el componente poligénico expresa el 19%, el segmento considerado del cromosoma 2 el 15% , y el del cromosoma 6 el 21%. En cuanto a las correlaciones genéticas, son relevantes los valores altamente positivos de los cromosomas 1, 3 y 6, mientras que, por el contrario, entre el cromosoma 2 y el componente poligénico las correlaciones genéticas son cercanas a 0. Estos resultados sugieren que la correlación genética entre ambos caracteres esta fuertemente determinada por el segmento del cromosoma 6.

Bibliografía

GEMAN S., GEMAN D., 1984. Stochastic relaxation, Gibbs distribution and the bayesian restorarion of images. IEEE. Trans. Pattern. Anal. Mach. Intell. 6:721-741.

GRINOLA F.E., HOESCHELE I., TIER B., 1996. Mapping quantitative trait loci in outcross populations via residual maximum likelihood. Genet. Sel. Evol. 28:479-400.

HALEY C.S., KNOTT S.A. ELSEN J.M., 1994. Mapping quantitative trait loci in crosses between outbred lines using least squares. Genetics: 136: 1195-1207.

LIU Z., DEKKERS J.C.M., 1998. Least squares interval mapping of quantitative trait loci under the infinitesimal genetic model in outbred populations. Genetics. 148: 495-505.

PÉREZ-ENCISO M., VARONA L., 2000. Quantitative Trait Loci Mapping in F2 crosses between outbred lines. Genetics 155.

PÉREZ-ENCISO M., VARONA L., ROTHSCHILD M., 2000. Computation of identity by descent probabilities conditional on DNA markers via a Monte Carlo Markov Chain method. Genet. Sel. Evol. (sometido).

VARONA L., GARCÍA CORTES L.A., PÉREZ-ENCISO, M. 2000. Bayes Factors for detection of quantitative trait loci. Genet. Sel. Evol. (sometido).

VARONA L., PÉREZ-ENCISO M., 1998. Detección de QTLs mediante la partición de la varianza genética en función del parentesco atribuible a segmentos del genoma. Información Técnica Económica Agrária. Vol 94ª Nº 3: 265-270.

VISSCHER P.M., HALEY C., 1996. Detection of putative quantitative trait loci in line crosses under infinitesimal genetic models. Theoretical Applied Genetics. 93: 691-702.

WANG C.S., RUTLEDGE J.J., GIANOLA D., 1994. Bayesian analysis of mixed model using Gibbs sampling with an application to litter size in Iberian Pigs. Genet. Sel. Evol. 26:91-115.

ITEA (2000), Vol. 96A N.° 3, 383-387

INFERENCIA EN POBLACIONES DONDE SE DESCONOCE LA CANTIDAD DE INFORMACIÓN PERDIDA

G. Yagüe-Utrilla, C. Moreno, L.A. García-Cortés, J. Altarriba

Unidad de Genética Cuantitativa y Mejora Animal, Facultad de Veterinaria, Universidad de Zaragoza, 50013 Zaragoza, España

RESUMEN

Se estudia la realización de inferencias en poblaciones con pérdida de información, dependiente de un proceso selectivo, en los casos en que se desconoce la magnitud de esa pérdida. Se han aplicado dos procedimientos: por un lado se modelizó directamente este proceso a través de la utilización del modelo truncado Tobit. Por otro, se analizaron los efectos de la inclusión del número de datos faltantes como una variable del modelo.

Introducción

Con el objeto de realizar inferencias correctas en una población sometida a un proceso selectivo que provoca una pérdida de información, suele requerirse la inclusión de dicho proceso en el modelo de estimación. Según la teoría de datos perdidos, "missing data theory" (RUBIN, 1976) una forma flexible de abordar este problema consiste en la introducción del proceso de pérdida de datos en la función de verosimilitud, idea que ya ha sido aplicada en el campo de la mejora genética animal por IM *et al.* (1989).

Bajo la teoría bayesiana, y en el supuesto de que la cantidad de información faltante sea conocida, es posible conseguir estimadores adecuados en estas poblaciones mediante el muestreo de Gibbs (GEMAN y GEMAN, 1984). Además, la inclusión en el modelo del vector de datos observados y perdidos como una variable aumentada (TANNER, 1993) facilita la implementación del muestreo de Gibbs. Esto se debe a que las condicionales que se obtienen con este procedimiento tienen una forma distribucional conocida (GELFAND, 1992; CHIB, 1992; SORENSEN *et al.*, 1998).

Sin embargo, la resolución de situaciones en las que sólo se dispone de los datos observados no es inmediata. Nuestro objetivo es realizar inferencias correctas bajo este supuesto, presentando a continuación las dos estrategias con las que hemos abordado este problema.

Descripción del modelo

Se utilizó un modelo animal:

$$y_i^* = \mu + a_i + e_i,$$

donde y_i^* son los datos en el caso de información completa, lo que implícitamente incluye la información genealógica. Consecuentemente, μ es la media general; a_i y e_i

son, respectivamente, el valor genético y el residuo asignados al individuo *i*. Por último, las distribuciones de las dos variables aleatorias son $a \sim N(0, A\sigma_a^2)$ y $e \sim N(0, I\sigma_e^2)$. A efectos de simplificar el análisis consideramos únicamente dos generaciones: generación parental (G_0) y generación filial (G_1). Sobre dicha población actúa un proceso de pérdida de información que queda reflejado en la variable z, de tal forma que:

$$z_i = \begin{cases} y_i^* \ si \ y_i^* \in G_0 \\ y_i^* \ si \ y_i^* \in G_1 \ , y_i^* \geq t \end{cases}$$

$$z_i = miss \ \ si \ y_i^* \in G_1, \ y_i^* < t$$

Analizamos la realización de inferencias en esta población bajo los siguientes procedimientos:

• El primer enfoque supone la modelización directa de la pérdida de información. El patrón de pérdida con el que tratamos es un truncamiento de los datos de la generación filial a partir de un umbral conocido *t*, por lo que esta modelización pasa por la aplicación del modelo Tobit (AMEMIYA, 1985).

• La alternativa a esta estrategia surge como consecuencia de la introducción del número de datos faltantes (*nf*) como una variable del modelo, lo que conduce a una distribución posterior de tipo mezcla (ROBERT, 1996).

Aplicación del modelo Tobit

Los datos observados junto con el patrón de pérdida de información componen nuestro conocimiento acerca de la población. Como ya se ha dicho, pensamos que sería posible modelizar directamente esta situación dado que conocemos el proceso de pérdida de información y que ésta se debe al truncamiento por la izquierda de los fenotipos de la generación filial a partir de un umbral conocido *t*.

En un supuesto de una distribución normal truncada por la izquierda, con una media μ y una varianza σ^2, la verosimilitud del modelo Tobit truncado (AMEMIYA, 1985) es:

$$L = \prod_{i=1,n} \sigma^{-1} \frac{exp\left[\frac{(y_i - \mu)^2}{\sigma^2}\right]}{1 - \Phi\left(\frac{t - \mu}{\sigma}\right)}$$

Aplicando este modelo a nuestro patrón de pérdida de información la distribución posterior de las variables será:

$$p\left(\mu, a, \sigma_a^2, \sigma_e^2 | z_{obs}\right) \propto p\left(z_{obs} | \mu, a, \sigma_a^2, \sigma_e^2\right)$$

$$p\left(a \mid A\sigma_a^2\right)$$

Cuadro 1

Resultados de las estimaciones de μ, σ_a^2 y σ_e^2 en tres casos de pérdida de información. Se muestra la media de las esperanzas obtenidas en las 15 repeticiones realizadas. Entre paréntesis se presentan la desviación típica de dichas esperanzas

	Valores verdaderos	Pérdida = 20%	Pérdida = 50%	Pérdida = 80%
μ	100	100.122 (0,242)	100.121 (0,181)	100.049 (0,140)
σ_a^2	10	10.630 (1,345)	6.339 (2,445)	4.154 (1,808)
σ_e^2	10	10.481 (1,400)	15.211 (2,163)	16.528 (1,885)

Donde z_{obs} se refiere a la información registrada. Esta expresión se divide en:

$$p\left(\mathbf{z}_{obs}\middle|\mu,\mathbf{a},\sigma_a^2,\sigma_e^2\right)\propto$$

$$\propto\prod_{i\in G0}\left\{\sigma_e^{-1/2}exp\left[\frac{\left(z_i-\mu-a_i\right)^2}{-2\sigma_e^2}\right]\right\}$$

$$\prod_{i\in G1}\left\{\frac{\sigma_e^{-1/2}exp\left[\frac{\left(z_i-\mu-a_i\right)^2}{-2\sigma_e^2}\right]}{1-\Phi\left(\frac{t-\mu-a_i}{\sigma_e}\right)}\right\}$$

y $p\left(\mathrm{a}\mid\sigma_a^2,\mathrm{A}\right)\propto\sigma_a^{-n/2}exp\left(\frac{\mathrm{a'\,Aa}}{-2\sigma_a^2}\right)$

Se decide llevar a cabo un muestreo de Gibbs para realizar inferencias sobre dicha población. Sin embargo, las condicionales obtenidas no se corresponden con distribuciones estándar, y por esta razón se ha utilizado un Metrópolis (TANNER, 1993) dentro de Gibbs para realizar el muestreo de valores.

A continuación se muestran los resultados de la aplicación de este modelo con distintos grados de pérdida de información. Se han considerado tres situaciones: pérdidas del 20%, 50%, 80%. Los parámetros verdaderos fueron en todas las ocasiones de: μ =100; σ_a^2 =10; σ_e^2 = 10.

Se destaca que con pérdidas leves de información este modelo ofrece resultados acordes con los valores simulados, sin embargo, no es capaz de ello en situaciones más restrictivas. Es necesario señalar que la condicional para los valores genéticos de los individuos de la generación parental no se ve modificada con respecto a la obtenida en el caso en el que no se tenga en cuenta la información perdida. Esto provoca que los hijos conocidos sean los únicos que aparezcan en dicha condicional, lo que parece indicar que de alguna manera no se tiene en cuenta toda la información perdida.

Utilización de una verosimilitud mezcla

Con el propósito de poner de manifiesto las dificultades que entraña este procedimiento, compararemos el caso de número de datos faltantes (*nf*) conocido (caso I), donde la dimensión del vector *z* es *n*, con el caso de *nf* desconocido (caso II) en el que no sabemos ni siquiera la dimensión de este vector, sino sólo la parte correspondiente a los datos observados.

a) Caso I:

En este caso tratamos con un modelo censurado en vez de truncado (Gauderman, 1994). Con el fin de simplificar el análisis aplicamos la técnica de la variable aumentada (TANNER, 1987), mediante la introducción de y^* en el modelo. Con ello se consiguen unas distribuciones condicionales de las que es sencillo muestrear (SORENSEN, 1998). La distribución posterior de los parámetros en el caso I es entonces:

$$p\left(\mu,\mathrm{a},\sigma_a^2,\sigma_e^2,\mathrm{y}^*\middle|\mathrm{z},nf\right)\propto p\left(\mathrm{z}\middle|\mathrm{y}^*\right)$$

$$p\left(\mathrm{y}^*\middle|\mu,\mathrm{a},\sigma_a^2,\sigma_e^2\right)\sum_{k=1}^{m}\left[w_k p_k\left(\mathrm{a}\mid\mathrm{A}_k\sigma_a^2\right)\right].$$

Cabe destacar que debido a la pérdida producida debemos tener en cuenta todas las genealogías posibles, lo que da como resultado una distribución tipo mezcla "*mixture*" expresada en el sumatorio de dichas genealogías (ROBERT, 1996). Siendo *m* el número total de genalogías distintas posibles, de tal forma que $w_k = 1/m$, dado que todas ellas son igualmente probables *a priori*.

Por otra parte,

$$p\left(y_i^* \middle| \mu, a_i, \sigma_a^2, \sigma_e^2, z_i = miss\right) \propto \Phi\left(\frac{t - \mu - a_i}{\sigma_e}\right)$$

es decir, una normal truncada, de la que es sencillo muestrear. El resto de las condicionales tiene formas distribucionales conocidas, y por ello, en esta situación, la inferencia se desarrolla satisfactoriamente (SORENSEN, 1998).

b) Caso II:

Sin embargo la distribución posterior se modifica cuando incluimos la cantidad de datos perdidos en el modelo en las ocasiones en que sólo se dispone de los datos registrados, dando lugar a:

$$p\left(\mu, a, \sigma_a^2, \sigma_e^2, y^*, nf|z\right) \propto \sum_{j=0}^{\infty} \left\{ p\left(nf_j\right) p\left(z|y^*, nf_j\right) \right.$$

$$\left. p\left(y^* \middle| \mu, a, \sigma_a^2, \sigma_e^2, nf_j\right) \sum_{k=1}^{m} \left[w_k p_k \left(a \mid A_k \sigma_a^2\right) \right] \right\}$$

Se aprecia que nos encontramos con una distribución tipo mezcla que afecta incluso a la verosimilitud de los datos, ya que la dimensión de z depende de *nf*. El análisis de este modelo debería seguir básicamente los mismos pasos que el anterior, sin em-bargo, la inclusión de *nf* en el modelo presenta problemas.

Una de las principales dificultades con las que nos hemos encontrado es la asignación de una distribución *a priori* para el número de datos faltantes. Si consideramos una *a priori* no informativa, la condicional resultante de *nf* no tiene máximo (figura 1), por lo que la cantidad perdida estimada siempre es igual a 0. Esto conduce a ignorar la pérdida de información, y como consecuencia, realizar inferencias incorrectas.

En casos sencillos se ha comprobado que la utilización de una *a priori* para *nf* a partir de la cuál se obtenga una distribución condicional binomial permite la obtención de resultados aceptables.

Por otro lado, el continuo cambio de dimensión del modelo que produce la modificación de *nf* induce a contemplar a las estrategias de elección de modelos como otra alternativa para abordar este tipo de problemas.

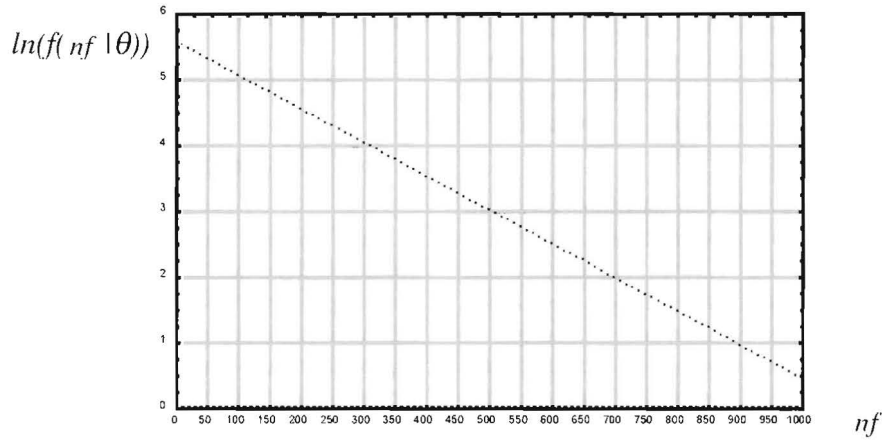


Figura 1. Forma distribucional del logaritmo de la condicional de datos faltantes (con la *a priori* no informativa)

Bibliografía

AMEMIYA T., 1984. Tobit models: a survey, Journal of Econometrics 24: 3.61.

CHIB S., 1992. Bayes inference in the Tobit censored regression model, Journal of Econometrics 51: 79-99.

GAUDERMAN W.J., THOMAS D.C., 1994. Censored survival models for genetic epidemiology: a Gibbs sampling approach, Genet. Epidemiol. 11: 171-188.

GELFAND A.E., SMITH F.M., LEE T., 1992. Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling, J. Am. Stat. Assoc. 87: 523-532.

GEMAN S., GEMAN D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Trans. Pattern Analysis and Machine Intelligence 6: 721-741.

IM S., FERNANDO R.L., GIANOLA D., 1989. Likelihood inferences in animal breeding under selection: a missing-data theory view point, Genet. Sel. Evol. 21: 399-414.

ROBERT C.P., 1996. Mixtures of distributions: inference and estimation, in Gilks W.R., Richardson S., Spiegelhalter D.J. (Eds.), Markov Chain Monte Carlo in practice, Chapman and Hall, London, pp. 419-440.

RUBIN D.B., 1976. Inference and missing data, Biometrika 63: 581-592.

SORENSEN D.A., GIANOLA D., KORSGAARD I., 1998. Bayesian mixed-effects model analysis of a censored normal distribution with animal breeding applications, Acta Agric. Scand., Sect. A, Animal Sci. 48: 222-229.

TANNER M.A., 1993. Tools for statistical inference, Springer-Verlag, Berlin.

TANNER M.A., WONG W.H., 1987. The calculation of posterior distributions by data augmentation, J. Am. Stat. Assoc. 82: 528-540.

CONCESIÓN DEL PREMIO PRENSA AGRARIA 1999 DE AIDA

Presidente
Leonardo Plana Claver

Vocales
Rafael Picorel Castaño
Miguel Valls Ortiz
Rafael Gella Fañanas

Secretario
Joaquín Uriarte Abad

Reunido el Jurado del Premio 1999 de Prensa Agraria de AIDA formado por D. Leonardo Plana Claver, Presidente de AIDA, D. Rafael Picorel Castaño, Director de la Estación Experimental de Aula Dei (CSIC), D. Miguel Valls Ortiz, Director del Instituto Mediterráneo de Zaragoza (IAMZ) y de Rafael Gella Fañanas, Jefe del Servicio de Investigación Agroalimentaria (DGA), actuando como Secretario D. Joaquín Uriarte Abad, Subdirector de ITEA, tal como establecen las bases de la convocatoria aprobadas en la Asamblea General de la Asociación celebrada en mayo de 1983, acordó premiar entre los artículos publicados en ITEA durante el año 1999 a los siguientes:

Serie Producción Vegetal: "Respuesta del sorgo al déficit hídrico en invernadero, I. Fenología, consumo de agua en los procesos de evapotranspiración y transpiración, y rendimientos", siendo sus autores: M.ª J. BERENGUER MERELO, J.M.ª FACI GONZÁLEZ.

Serie Producción Animal: "Diferencias genéticas entre grupos de importación en cuatro líneas de porcino españolas", siendo sus autores: D. BABOT, J.L. NOGUERA, L. ALFONSO, J. ESTANY.

Zaragoza, 24 de mayo de 2000

PREMIOS DE PRENSA AGRARIA 2000 DE LA ASOCIACIÓN INTERPROFESIONAL PARA EL DESARROLLO AGRARIO

La Asociación Interprofesional para el Desarrollo Agrario (AIDA) acordó en Asamblea General celebrada en mayo de 1983, instaurar un premio anual de Prensa Agraria, con el objetivo de hacer destacar aquel artículo de los publicados en ITEA que reúna las mejores características técnicas, científicas y de valor divulgativo, y que refleje a juicio del jurado, el espíritu fundacional de AIDA de hacer de transmisor de conocimientos hacia el profesional, técnico o empresario agrario.

El día 9 de abril de 1987, la Junta Directiva de AIDA aceptando la propuesta del Jurado del Premio ITEA 1986 instituyó dos premios; uno para los artículos publicados en la sección de Producción Animal y otro para aquellos que aparezcan en la sección de Producción Vegetal.

Los premios se regirán de acuerdo a las siguientes

BASES

1. Podran concursar todos los artículos que versen sobre cualquier tema técnico-económico-agrario.
2. Los artículos que podrán acceder a los premios serán todos aquellos que se publiquen en ITEA en el año 2000. Consecuentemente, los originales deberán ser enviados de acuerdo con las normas de ITEA y aprobados por su Comité de Redacción.
3. El jurado estará constituido por las siguientes personas:
   a) Presidente de AIDA, que presidirá el jurado.
   b) Director de la revista ITEA, que actuará de Secretario.
   c) Jefe del Servicio de Investigación Agroalimentaria de Zaragoza (Diputación General de Aragón).
   d) Director del Instituto Agronómico Mediterráneo de Zaragoza.
   e) Director de la Estación Experimental de Aula Dei.
4. Los premios serán anuales y con una dotación de 50.000 ptas. cada uno.
5. Las deliberaciones del jurado serán secretas, y su fallo inapelable.
6. El fallo del jurado se dará a conocer en la revista ITEA, y la entrega del premio se realizarán con motivo de la celebración de las Jornadas de Estudio de AIDA.

CIHEAM

CENTRO INTERNACIONAL DE ALTOS ESTUDIOS AGRONÓMICOS MEDITERRÁNEOS
INSTITUTO AGRONÓMICO MEDITERRÁNEO DE ZARAGOZA

CIHEAM/IAMZ - Cursos 1998-99-00

	CURSOS	FECHAS	LUGAR	ORGANIZACIÓN
PRODUCCIÓN VEGETAL	*MEJORA GENÉTICA VEGETAL	5 Oct. 98/ 11 Jun. 99	Zaragoza	IAMZ
	*OLIVICULTURA Y ELAIOTECNIA	1 Oct. 99/ 26 Mayo 00	Córdoba	UCO/CA-JA/CSIC/ INIA/COI/IAMZ
	UTILIZACIÓN SOSTENIBLE DE RECURSOS EN LOS SISTEMAS DE AGRICULTURA DE SECANO DE LA REGIÓN MEDITERRÁNEA	18-29 Oct. 99	Rabat	IAMZ/IAV Hassan II
	INTRODUCCIÓN DE PLANTAS TRANSGÉNICAS EN LA AGRICULTURA: EVALUACIÓN Y CRITERIOS DE DECISIÓN	22-26 Nov. 99	Zaragoza	IAMZ
	USO DE LOS MARCADORES MOLECULARES EN MEJORA VEGETAL	7-18 Feb. 00	Barcelona	IAMZ/IRTA
	OPTIMIZACIÓN DE LOS PROCESOS DE MEJORA DE ESPECIES LEÑOSAS FRUTALES Y FORESTALES: ASPECTOS FISIOLÓGICOS Y GENÉTICOS	6-10 Mar. 00	Zaragoza	IAMZ
	OPTIMIZACIÓN DE LOS PROCESOS DE SELECCIÓN Y MEJORA DE ESPECIES LEÑOSAS FRUTALES Y FORESTALES: DISEÑO DE EXPERIMENTOS Y ANÁLISIS ESTADÍSTICO	13-17 Mar. 00	Zaragoza	IAMZ
	CONTROL INTEGRADO DE PLAGAS Y ENFERMEDADES EN CULTIVOS PROTEGIDOS	5-16 Jun. 00	Zaragoza	IAMZ
PRODUCCIÓN ANIMAL	REQUISITOS DE CALIDAD DE LA CANAL Y DE LA CARNE DE RUMIANTES PARA SU COMERCIALIZACIÓN	19-30 Oct. 98	Zaragoza	IAMZ
	PLANIFICACIÓN Y DESARROLLO DE CAMPAÑAS DE SANEAMIENTO GANADERO	8-12 Feb. 99	Zaragoza	IAMZ/FAO
	TÉCNICAS MOLECULARES EN MEJORA GENÉTICA ANIMAL	15-26 Mar. 99	León	IAMZ/Univ. León
	PASTORALISMO Y DESARROLLO EN ZONAS ÁRIDAS Y SEMIÁRIDAS MEDITERRÁNEAS	22 Mar./ 3 Abr. 99	Rabat	IAMM/IAMZ/ IAV Hassam II
	APLICACIONES DE LA TECNOLOGIA NIRS EN LA EVALUACIÓN DE PRODUCTOS AGRARIOS	12-16 Abr. 99	Zaragoza	IAMZ/UCO
	DIVERSIFICACIÓN DE ESPECIES DE PECES EN LA ACUICULTURA MARINA MEDITERRÁNEA	24-28 Mayo 99	Zaragoza	IAMZ/FAO
	*PRODUCCIÓN ANIMAL	4 Oct. 99/ 9 Jun. 00	Zaragoza	IAMZ
	*ACUICULTURA	10 Ene./ 30 Jun. 00	Las Palmas de Gran Canaria	ULPGC/ICCM/ IAMZ
	VALORIZACIÓN DE PRODUCTOS LÁCTEOS DE OVINOS Y CAPRINOS EN EL ÁREA MEDITERRÁNEA. TECNOLOGIAS ACTUALES Y PERSPECTIVAS DE MERCADO	10-19 Abr. 00	Surgères	IAMZ/ENILIA

	CURSOS	FECHAS	LUGAR	ORGANIZACIÓN
MEDIO AMBIENTE	RÍOS Y RIBERAS DE RÉGIMEN MEDITERRÁNEO Y SU GESTIÓN	21 Sep./ 2 Oct. 98	Zaragoza	IAMZ
	ARBUSTOS FORRAJEROS: SU PAPEL EN EL DESARROLLO Y LA CONSERVACIÓN DEL MEDIO AMBIENTE DE LAS ZONAS ÁRIDAS Y SEMIÁRIDAS MEDITERRÁNEAS	28 Sep./ 9 Oct. 98	Rabat	IAMZ/IAV Hassan II
	*ORDENACIÓN RURAL EN FUNCIÓN DEL MEDIO AMBIENTE	5 Oct. 98/ 11 Jun. 99	Zaragoza	IAMZ
	ECONOMÍA DE LOS RECURSOS NATURALES	15-26 Mar. 99	Zaragoza	IAMZ
	DEFENSA CONTRA INCENDIOS FORESTALES	18-29 Oct. 99	Zaragoza	IAMZ/MIMAM/FAO
	USO CONJUNTO DE AGUAS SUPERFICIALES Y SUBTERRÁNEAS	8-12 Nov. 99	Zaragoza	IAMZ
	EVALUACIÓN DEL IMPACTO AMBIENTAL DE LAS INSTALACIONES ACUÍCOLAS EN EL MEDITERRÁNEO	17-21 Ene. 00	Zaragoza	IAMZ/FAO
	ESTRATEGIAS DE REPOBLACIÓN FORESTAL EN LA REGIÓN MEDITERRÁNEA	7-18 Feb. 00	Zaragoza	IAMZ
	EVALUACIÓN Y MANEJO DE RECURSOS PESQUEROS EN EL MEDITERRÁNEO	27-31 Mar. 00	Fuengirola	IAMZ/IEO
	TURISMO EN ZONAS RURALES: ESTRATEGIAS Y PROMOCIÓN	3-7 Abr. 00	Zaragoza	IAMZ
	ESTADO ECOLÓGICO DE LAS AGUAS SUPERFICIALES: MÉTODOS DE MEDIDA Y ESTRATEGIAS DE GESTIÓN	8-19 Mayo 00	Zaragoza	IAMZ
COMERCIALIZACIÓN	GESTIÓN DE LA CALIDAD. NUEVOS CONCEPTOS Y SU APLICACIÓN EN EL MARKETING AGROALIMENTARIO	9-20 Nov. 98	Zaragoza	IAMZ
	COMERCIALIZACIÓN DE LOS PRODUCTOS DEL MAR: TENDENCIAS Y RETOS	14-18 Dic. 98	Zaragoza	IAMZ
	GESTIÓN DE LOS CANALES DE DISTRIBUCIÓN ALIMENTARIA	11-22 Ene. 99	Zaragoza	IAMZ
	LA ORGANIZACIÓN MUNDIAL DEL COMERCIO Y SU IMPACTO EN EL MARKETING INTERNACIONAL AGROALIMENTARIO	22-26 Feb. 99	Zaragoza	IAMZ/OMC
	DESARROLLO DE NUEVOS PRODUCTOS EN EL SECTOR AGROALIMENTARIO	26 Abr./ 7 Mayo 99	Zaragoza	IAMZ
	*MARKETING AGROALIMENTARIO	4 Oct. 99/ 9 Jun. 00	Zaragoza	IAMZ

(*) Cursos de Especialización Postuniversitaria del correspondiente Programa Master of Science. Se desarrollan cada dos años:

- MEJORA GENÉTICA VEGETAL: 98-99; 00-01; 02-03
- OLIVICULTURA Y ELAIOTECNIA: 99-00; 01-02; 03-04
- PRODUCCIÓN ANIMAL: 99-00; 01-02; 03-04
- ACUICULTURA: 99-00; 01-02; 03-04
- ORDENACIÓN RURAL EN FUNCIÓN DEL MEDIO AMBIENTE: 98-99; 00-01; 02-03
- MARKETING AGROALIMENTARIO: 99-00; 01-02; 03-04

Se destinan primordialmente a titulados superiores en vías de especialización postuniversitaria. No obstante se estructuran en ciclos independientes para facilitar la asistencia de profesionales interesados en aspectos parciales del programa. Los participantes que cumplan los requisitos académicos pueden optar a la realización del 2º año para la obtención del Título Master of Science. El plazo de inscripción para los cursos de Producción Animal, Marketing Agroalimentario y Olivicultura y Elaiotecnia finaliza el 15 de Mayo 1999. El plazo de inscripción para el curso de Acuicultura finaliza el 30 de Junio 1999. El plazo de inscripción para los cursos de Mejora Genética Vegetal y Ordenación Rural en Función del Medio Ambiente finaliza el 15 de Mayo 2000.

Los cursos de corta duración están orientados preferentemente a investigadores y profesionales relacionados en el desarrollo de sus funciones con la temática de los distintos cursos. El plazo de inscripción para los cursos de corta duración finaliza 90 días antes de la fecha de inicio del curso.

Los candidatos de países miembros del CIHEAM (Albania, Argelia, Egipto, España, Francia, Grecia, Italia, Líbano, Malta, Marruecos, Portugal, Túnez y Turquía) podrán solicitar becas que cubran los derechos de inscripción, así como becas que cubran los gastos de viaje y de estancia durante el curso. Los candidatos de otros países interesados en disponer de financiación deberán solicitarla directamente a otras instituciones nacionales o internacionales.

En la página Web se proporciona información sobre los cursos programados y se facilita el formulario de inscripción: http://www.iamz.ciheam.org

Para mayor información dirigirse a:

Instituto Agronómico Mediterráneo de Zaragoza
Apartado 202 - 50080 ZARAGOZA (ESPAÑA)
Teléfono 976 57 60 13 - Fax 976 57 63 77
e-mail: iamz@iamz.ciheam.org – Web http://www.iamz.ciheam.org

INSCRIPCIÓN EN AIDA

* Si desea Ud. pertenecer a la Asociación, rellene la ficha de inscripción así como la carta para la domiciliación del pago de la cuota de asociado y envíelas a AIDA. Aptdo. 727. 50080 Zaragoza.

El abajo firmante solicita su inscripción como miembro de la Asociación Interprofesional para el Desarrollo Agrario.

Apellidos ..... Nombre .....

Dirección postal .....

Teléfono .....

Profesión..... Empresa de trabajo .....

Área en que desarrolla su actividad profesional .....

CUOTA ANUAL: Firma.

❑ Sólo una Serie de ITEA { ❑ P. Animal / ❑ P. Vegetal } 4.000 ptas. ó 24 €

❑ Ambas Series 5.500 ptas. ó 33 €

FORMA DE PAGO:

❑ Cargo a cuenta corriente o libreta
❑ Cheque bancario
❑ Cargo a tarjeta
❑ VISA
❑ MASTERCARD

Tarjeta número:

□□ □□□□□□□□□□□□□□□□ Fecha de caducidad: ..... / .....

SR. DIRECTOR DE .....

Muy Sr. mío:

Ruego a Vd. se sirva adeudar en la cuenta cte./libreta n.º ..... que matengo en esa oficina, el recibo anual que será presentado por la "Asociación Interprofesional para el Desarrollo Agrario".

Atentamente,

Firmado:

BANCO O CAJA DE AHORROS: .....

SUCURSAL: .....

DIRECCIÓN CALLE/PLAZA: ..... N.º .....

CÓDIGO POSTAL: .....

POBLACIÓN: .....

INFORMACIÓN PARA AUTORES

Tipo de artículos que pueden ser enviados para su consideración al Comité de Redacción: se admite todo aquel que contribuya al intercambio de información profesional y trate de los más recientes avances que existan en las distintas actividades agrarias.

Una *información para autores* más detallada puede ser solicitada al Comité de Redacción. Rogamos sea leída detenidamente, prestando atención especial a los siguientes puntos:

CONDICIONES GENERALES

Los artículos, en castellano, serán enviados por triplicado a:
Sr. Director de la Revista ITEA - Apartado 727 - 50080 ZARAGOZA

RECOMENDACIONES EN LA PREPARACIÓN DE LOS ORIGINALES

La extensión máxima será de 25 folios de texto mecanografiado a doble espacio, cuadros y figuras incluidos. Los artículos que superen dicha extensión serán considerados sólo excepcionalmente.

Los artículos se remitirán a dos evaluadores anónimos expertos en el tema y el autor recibirá un informe del Comité de Redacción con las correcciones de dichos evaluadores. Una vez realizadas las correcciones, el autor enviará un sólo ejemplar mecanografiado y una copia en disquete, para agilizar el trabajo en imprenta. Si el Comité de Redacción considera que se han atendido las consideraciones del informe, enviará una carta de aceptación al remitente, y el artículo pasará de inmediato a imprenta.

Los autores recibirán un juego de las primeras pruebas de impresión que deberán ser revisadas y devueltas rápidamente a la Redacción. El retraso en el retorno de las pruebas determinará que el artículo sea publicado con las correcciones del Comité de Redacción.

El título no incluirá abreviaturas y será corto y preciso. En la misma página se incluirán los nombres completos de los autores, así como la dirección postal y nombre de la Entidad en donde se haya realizado el trabajo.

Se incluirá en primer lugar un resumen corto de 200-250 palabras y hasta seis palabras clave. Además, se añadirá un resumen en *inglés* de la misma extensión, sin olvidar el *título* traducido y las palabras clave (Keywords).

A continuación del resumen vendrá el artículo completo, procurando mantener una disposición lógica, considerando cuidadosamente la jerarquía de títulos, subtítulos y apartados.

Los dibujos, gráficos, mapas y fotografías deben titularse todos *figuras*. Los *cuadros* y *figuras* deben llevar numeración diferente, pero ambos en cifras árabes.

Los pies o títulos de cuadros y figuras deben redactarse de modo que el sentido de éstos pueda comprenderse sin necesidad de acudir al texto. Los títulos, pies y leyendas de los cuadros y figuras se traducirán al inglés y se incluirán en letra cursiva, bajo el correspondiente en español.

Los dibujos, gráficos, mapas, fotografías y diapositivas serán presentados en la mejor calidad posible.

En general se evitará el uso de abreviaturas poco conocidas, que en todo caso serán debidamente explicadas.

REFERENCIAS BIBLIOGRÁFICAS

En el *texto* las referencias deben hacerse mediante el apellido de los autores en mayúsculas seguido del año de publicación, todo entre paréntesis.

Al final del trabajo, y precedida de la mención *Referencias Bibliográficas,* se hará constar una lista alfabética de todas (y únicamente) las referencias utilizadas en el texto. En el caso de incluir varios trabajos del mismo autor se ordenarán cronológicamente.

Cuando se citen revistas[1], libros[2], capítulos de libro[3] y comunicaciones a congresos[4] se hará según los siguientes ejemplos:

(1) HERRERO J., TABUENCA, M.C., 1966. Épocas de floración de variedades de hueso y pepita. An. Aula Dei, 8 (1), 154-167.

(2) STELL, R.G.D., Y TORRIE, J.H. 1986. Bioestadística: principios y procedimientos (segunda edición) 622 pp. Ed. McGraw-Hill. México.

(3) GAMBORG O.L., 1984. Plant cell cultures: nutrition and media, pp. 18-26. En: Cell Culture and Somatic Cell Genetics of Plants. Vol. 1, I.K. Vasil (Ed.), 825 pp. Ed. Academic Press, Orlando (EEUU).

(4) ANGEL I., 1972. The use of fasciculate form (determinate habit) in the breeding of new Hungarian pepper varieties. Third Eucarpia Meeting on Genetics and Breeding of Capsicum, 17-24, Universidad de Turín (Italia).