

¿Pueden usarse animales cruzados en selección genómica?

N. Ibáñez-Escriche*, R.L. Fernando**, J.C.M. Dekkers**

* Genética y Mejora Animal. IRTA, Lleida, 25191. Spain.

** Department of Animal science, Iowa State University, Ames, Iowa 50011

E-mail: Noelia.ibanez@irta.es

Resumen

La producción en poblaciones “puras” suele tener una baja reproducibilidad en sus descendientes “cruzados”. La selección genómica podría utilizarse para evaluar poblaciones “puras” usando los datos de sus descendientes “cruzados”. Sin embargo, en las poblaciones cruzadas quizás el desequilibrio de ligamiento (LD) no está restringido a marcadores estrechamente ligados al QTL y los efectos de los marcadores podrían ser específicos de cada población. Estos dos problemas podrían solucionarse utilizando un modelo con los alelos de los SNPs específicos para cada población. Para investigar esta idea usamos un modelo con los efectos de los genotipos de los SNPs (modelo 1) y otro modelo con los efectos de alelos de los SNPs específicos para cada población (modelo 2). Ambos modelos se utilizaron para predecir los valores genéticos de las poblaciones “puras” usando datos F_1 . Tres situaciones fueron simuladas, en las dos primeras se consideró que las dos poblaciones tenían un mismo origen con una diferencia de 50 y 550 generaciones, respectivamente. En la tercera situación se consideró que las dos poblaciones tenían orígenes distintos. En todos los casos las dos poblaciones generaron una población F_1 con un tamaño de 1.000 individuos. Los valores fenotípicos de la F_1 fueron simulados con una media de 12 QTL segregando y una heredabilidad de 0.3. En el análisis de la F_1 y la población “pura” de validación se escogieron 500 marcadores en segregación. Para estimar el efecto de los SNPs se utilizó el método Bayesiano llamado Bayes-B. La precisión media de los valores genéticos obtenida varió entre 0.789 y 0.718. Sin embargo, se observó que conforme las poblaciones estuvieron más alejadas la precisión disminuyó y el modelo 2 dio valores ligeramente superiores que el modelo 1. Estos resultados sugerirían que los animales cruzados pueden ser utilizados para evaluar poblaciones “puras”. Además modelos con origen específico de población darían mejores resultados.

Palabras clave: Selección genómica, Evaluación genética, Animales cruzados, Valores genéticos.

Summary

¿Can crossbred animals be used for genomic selection?

Performance of purebred parents can be a poor predictor of performance of their crossbred descendants. However, in crossbred populations linkage disequilibrium may not be restricted to markers that are tightly linked to the QTL and the effects of SNPs may be breed specific. Both these problems can be addressed by using a model with breed-specific SNP effects. To investigate this idea, we used a model with effects of SNP genotypes (model 1) and a model with breed-specific effects of SNP alleles (model 2) to predict purebred breeding values using F_1 data. Three scenarios were considered. In the first two, pure breeds were assumed to have a common origin either 50 or 550 generations ago. In the third scenario, the two breeds did not have a common origin. In all these scenarios, the two breeds were used to generate an F_1 with 1,000 individuals. Trait phenotypic values controlled by 12 segregating QTL and with a heritability of 0.30 were simulated for the F_1 individuals. Further, 500 segregating markers on a chromosome of 1 Morgan were chosen for analysis in the F_1 s and in the validation population of purebred. A Bayesian method (Bayes-B) was used to estimate the SNP effects. The accuracy of the predictions was between 0.789 and 0.718. However, the accuracy was

lower when the populations were more separate and model 2 gave values slightly higher than model 1. These results suggest that crossbred data could be used to evaluate purebreds and breed specific models could give better results.

Key words: Genomic selection, Genetic evaluation, Crossbred, Breeding values

Introducción

Es bien conocido que la producción de las poblaciones "puras" tiene una baja reproducibilidad en las poblaciones "cruzadas" comerciales (Dekkers et al., 2007). Para resolver este problema se han desarrollado métodos de evaluación de poblaciones "puras" usando datos de sus descendientes cruzados (Lo et al., 1993, Lo et al., 1997). Asimismo, hay estudios que demuestran que estos métodos de evaluación podrían incrementar la respuesta a la selección respecto a los métodos clásicos (Bijma y Van Arendonk, 1998). Sin embargo, su implementación en los sistemas de producción ha sido escasa debido principalmente al coste y dificultad que supone la trazabilidad de los datos tanto productivos como genealógicos de las poblaciones comerciales (Dekkers et al., 2007). Por otro parte, en producción animal la selección genómica está tomando una mayor relevancia a medida que se dispone de una mayor información genómica (p.ej. mayor número de SNPs). La selección genómica predice los valores genéticos totales (GEBVs) dada una alta densidad de marcadores a lo largo de todo el genoma (Meuwissen et al., 2001). Este método tiene dos ventajas substanciales respecto a los métodos de evaluación clásicos como el BLUP (Henderson, 1973). Por una parte, no requiere de genealogía y, por otra, una vez estimados los efectos de los marcadores permite predecir el valor genético sobre varias generaciones sin utilizar sus valores fenotípicos. No obstante, el éxito de esta metodología depende fundamentalmente de la

precisión en la predicción de los GEBVs. Varios autores han estudiado la potencia de esta precisión usando simulaciones por ordenador (Meuwissen et al., 2001; Habier et al., 2007; Calus et al., 2008). Sin embargo, estos estudios siempre han estado focalizados en poblaciones "puras". Las principales razones de este hecho son dos: 1) en las poblaciones "cruzadas" el desequilibrio de ligamiento (LD) podría no estar restringido a marcadores estrechamente ligados al QTL (Zhao et al., 2003), y 2) los efectos de los marcadores podrían ser específicos para cada población. No obstante, ambos problemas podrían solucionarse utilizando un modelo con los alelos de los SNPs específicos para cada población. El objetivo de este estudio fue investigar esta idea, comparando dos modelos que permiten predecir los valores genéticos de la población "pura" usando datos F_1 . En el primer modelo solo se tuvo en cuenta los genotipos de los SNPs y en el segundo modelo los alelos de los SNPs fueron específicos para cada población.

Material y Métodos

Modelo estadístico

Modelo 1

$$y = 1\mu + \sum_k x_k B_k \delta_k + e, \quad (1)$$

Modelo 2

$$y = 1\mu + \sum_{k_l} x_{k_l} B_{k_l} \delta_{k_l} + e, \quad (j=1 \text{ o } 2, l=1,2) \quad (2)$$

donde y es el vector de los datos fenotípicos, μ es la media, x_k y x_{k1j} son, respectivamente, los vectores columna correspondientes a los genotipos de los marcadores en el locus k (modelo 1) y a los alelos l de los marcadores en el locus k y en la población j (modelo 2), B_k y B_{k1j} son los efectos del marcador y los efectos de sus alelos, σ_k y σ_{k1j} son las variables indicadoras 0/1 para cada uno de los modelos y e es el vector de los efectos residuales. En x_k , el marcador de un genotipo de un individuo se codifica como el número de copias que tiene un SNP de un alelo, que en este caso puede ser 0, 1, ó 2.

Para estudiar el efecto de los marcadores (SNP) en cada modelo usamos el método Bayes-B descrito por Meuwissen et al. (2001). En Bayes-B la distribución de los datos, dados los demás parámetros, es una normal con media 1μ y varianza $(X\sigma_B^2X'+I\sigma_e^2)$. X es la matriz que contiene los vectores x_k o x_{k1j} , $\sigma_B^2 = \text{diag}\{\sigma_{Bk}^2\}$ o $\text{diag}\{\sigma_{Bklj}^2\}$ (modelo 2), σ_B^2 con-

tiene las varianzas de los efectos de los marcadores o alelos en los cuales σ_k o $\sigma_{k1j} = 1$. Para todos los parámetros desconocidos se asumieron las mismas distribuciones a priori que Meuwissen et al. (2001) en Bayes-B. Al igual que estos autores, la probabilidad de prior de que σ_{Bk}^2 ó σ_{Bklj}^2 no sea cero es la esperanza de la proporción de los QTL segregando del número total de QTL que están en el genoma. Nótese que cuando la σ_{Bk}^2 ó $\sigma_{Bklj}^2 = 0$ ($l =$ alelo 1, 2) en una iteración, σ_k o $\sigma_{k1j} = 0$ y el marcador o alelo (modelo 2) de ese locus no se incluye en esa iteración. Más detalles pueden encontrarse en Meuwissen et al. (2001).

Simulación

Tres situaciones diferentes, con similar estructura, fueron simuladas en este estudio (figura 1). En las dos primeras se asumió que las poblaciones A y B provenían de un ori-

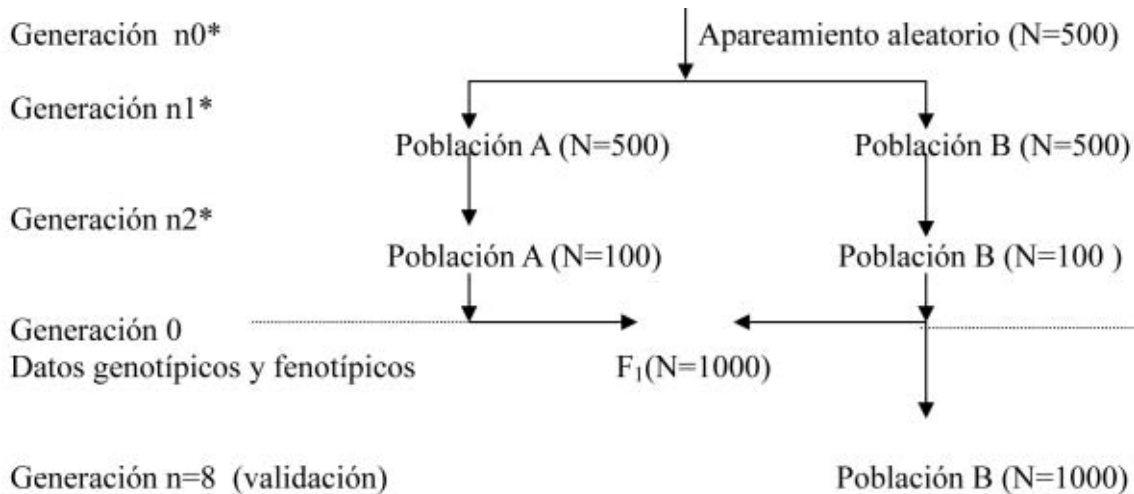


Figura 1. Simulación de las poblaciones.
Figure 1. Simulated populations.

* Situación 1: generación de separación de poblaciones y reducción n0= -1050, n1= n2= -50. Situación 2: n0= -1050, n1= -550, n2= -50. Situación 3: poblaciones (A, B) totalmente separadas n0= 0, n1= -1050, n2= -50.

gen común, con una diferencia de 50 y 550 generaciones de distancia, respectivamente. En la tercera situación se consideraron poblaciones con diferentes orígenes. En los tres casos, las poblaciones A y B fueron usadas para generar 1.000 individuos F_1 , donde se simularon individuos con valores fenotípicos controlados por una media de 12 QTL y una heredabilidad de 0.3. Además, para el análisis de la población F_1 y la de validación (figura 1), se simuló un solo cromosoma de donde se escogieron 500 marcadores segregando en 1 Morgan. Para la estimación de los GEBVs, dados los marcadores y valores fenotípicos, se utilizó la población F_1 , mientras que la población B se consideró para validar la precisión de los GEBVs y comparar los dos modelos. El criterio para comparar ambos modelos fue la correlación entre el verdadero valor genético y el estimado en la población B con 8 generaciones de distancia de la población F_1 . Esta correlación fue también definida como la precisión de los valores genéticos estimados. En todas las simulaciones se utilizó una distribución *gamma* para muestrear los efectos de los QTL bialélicos y una distribución *Bernoulli* con frecuencia 0.5 para sus alelos. En todas las generaciones se consideró una tasa de mutación de 0.000025. Las recombinaciones a lo largo del cromosoma fueron modeladas de acuerdo a una función binomial de mapeo.

Resultados y discusión

Los resultados del estudio corresponden a 20 réplicas. Para estimar los parámetros desconocidos de cada modelo, en cada réplica, se utilizó Bayes-B con una cadena de MCM de 10.000 iteraciones y un período de quemado de 1.000.

La tabla 1 muestra la precisión de los GEBVs estimados para la población B en las tres situaciones simuladas, así como la diferencia de precisión entre el modelo 2 y modelo 1. La precisión media obtenida varía entre 0.789 y 0.718, siendo similar a la obtenida en otros trabajos donde los animales de la población pura eran los utilizados para estimar los efectos de los marcadores (Meuwissen et al., 2003; Habier et al., 2007). Como se puede observar, conforme las poblaciones están más distanciadas la precisión es menor en ambos modelos. Esto podría deberse a que conforme las poblaciones están más alejadas su patrón de LD difiere en mayor medida, por lo que una mayor distancia entre poblaciones afectaría a la precisión de los GEBVs. Otro resultado relevante de este estudio (tabla 1) es que conforme las poblaciones están más alejadas (situación 2 y 3) el modelo 2 da una mayor precisión media que el modelo 1. Sin embargo, aunque la diferencia entre modelos es significativa, su relevancia es menor de la esperada dada la

Tabla 1. Precisión de los GEBVs para la población B de validación y la diferencia de precisión entre el modelo 2 y 1 (500 SNPs y ~ 12 QTL segregando)

Table 1. Accuracies of GEBVs for the validation population B and the difference of accuracy between model 2 and model 1 (500 SNPs and ~ 12 QTL segregating)

	Situación 1		Situación 2		Situación 3	
	Media	S.E	Media	S.E	Media	S.E
Modelo 1	0.778	0.014	0.723	0.047	0.718	0.034
Modelo 2	0.789	0.018	0.764	0.036	0.733	0.033
Diferencia	0.011	0.010	0.041	0.017	0.015	0.008

similitud de los valores en los dos modelos. Una de las posibles explicaciones de este hecho sería que, cuando las poblaciones están muy alejadas, gran parte de los marcadores que están segregando en la población F_1 se encuentran fijados en una de las poblaciones "puras", por lo que en estos casos los modelos 1 y 2 serían equivalentes.

Los resultados obtenidos en este estudio indicarían que los animales cruzados pueden ser utilizados para evaluar poblaciones "puras". Además, modelos con origen específico de población darían mejores resultados. Sin embargo estos resultados son preliminares y están basados en un modelo meramente aditivo y en una población F_1 . Sería interesante extender estas simulaciones a diferentes poblaciones (p.ej. F_2) y con modelos donde se tenga en cuenta los efectos dominantes.

Agradecimientos

Este trabajo ha sido realizado durante la estancia de investigación de la Dra. Noelia Ibáñez-Escriche en el *Department of Animal Science, Iowa State University (USA)* financiada por el Ministerio de Educación y Ciencia a través del programa José Castillejo (JC2007-00169).

Bibliografía

Bijma P, van Arendonk JAM, 1998. Maximizing genetic gain for the sire line of a crossbreeding

scheme utilizing both purebred and crossbred information. *J. Anim. Sci.*, 66, 529-542.

Calus MPL, Meuwissen THE, de Roos APW, Veerkamp RF, 2008. Accuracy of Genomic Selection Using Different Methods to Define Haplotypes. *Genetics*, 178, 553-561.

Dekkers JCM, 2007. Marker-assisted selection for commercial crossbred performance. *J. Anim. Sci.*, 85, 2104-2114.

Habier D, Fernando RL, Dekkers JCM, 2007. The impact of Genetic Relationship on Genome-Assisted Breeding Values. *Genetics*, 117, 2389-2397.

Henderson CR, 1973. Sire evaluation and genetic trends. In: *Animal breeding and genetics symposium in Honor of Dr. Jay L. Lush*, Champaign, IL. p 10-41.

Lo LL, Fernando RL, Grossman M, 1993. Covariance between relatives in multibreed populations - additive-model. *Theor. Appl. Genet.*, 87, 423-430.

Lo LL, Fernando RL, Grossman M, 1997. Genetic evaluation by blup in two-breed terminal crossbreeding systems under dominance. *J. Anim. Sci.*, 75, 2877-2884.

Meuwissen THE, Hayes BJ, Goddard ME, 2001. Prediction of Total Genetic Values Using Genome-Wide Dense Markers Maps. *Genetics*, 157, 1819-1829.

Zhao H, Nettlwtton D, Soller M, Dekkers JCM. 2005. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genet. Res.*, 86, 77-87.

(Aceptado para publicación el 28 de abril de 2008)