

POTENCIA ESTADÍSTICA DEL ANÁLISIS BAYESIANO DE LA DISTORSIÓN DE LA SEGREGACIÓN

Id-Lahoucine¹, S., Cánovas², A., Jatón^{2,3}, C., Sargolzaei^{2,3}, M. y Casellas¹, J.

¹Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), España.

²Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Canada.

³Semex Alliance, Guelph, ON, Canada
Samir.IdLahoucine@uab.cat

INTRODUCCIÓN

La distorsión de la segregación (DS) definida como la desviación en las frecuencias genotípicas esperadas en la descendencia (Silver, 1993; Pardo-Manuel de Villena et al., 2000) continúa siendo poco estudiada, especialmente en las especies ganaderas. La DS puede ser el resultado de muchos mecanismos biológicos distintos, afectando a gametos, embriones, fetos (Huang et al., 2013), o incluso durante la etapa postnatal (Moore, 2006). En este sentido, el estudio de la DS puede conducir a la identificación de factores genéticos implicados en la fertilidad y los caracteres reproductivos. El modelo desarrollado por Casellas et al. (2014) presenta una metodología aplicable a las poblaciones de ganado con una parametrización flexible de la DS y para todo tipo de genealogías. El objetivo principal de este estudio fue evaluar la potencia estadística para detectar la DS bajo distintas estructuras poblacionales y genéticas para posterior aplicación utilizando datos reales en distintas poblaciones de vacuno de leche y carne.

MATERIAL Y MÉTODOS

Se simularon generaciones parentales con un único marcador genético bialélico, asumiendo una frecuencia alélica entre 0,05 y 0,95. A partir de estos, se simularon diferentes tamaños poblacionales (N) con 25, 100, 200 y 500 individuos generados por apareamiento aleatorio bajo una estructura de 1, 25 o 100 hembras por semental (♀:♂). Los genotipos de la generación descendiente fueron simulados con diferentes niveles de DS entre -0,5 y 0,5 (modelización de la DS más abajo). Para cada escenario se generaron 10.000 réplicas.

Para la evaluación de la DS se utilizaron los modelos desarrollados por Casellas et al. (2014). A modo de ejemplo esta parametrización modela la probabilidad de heredar un alelo específico a partir de un progenitor heterocigoto (A/B) incluyendo el efecto de DS (α) como:

$$p(A)=1-p(B)=0,5+\alpha \text{ y } p(B)=1-p(A)=0,5-\alpha,$$

En el caso de los progenitores homocigotos se asume $\alpha=0$. Alternativamente, la DS se puede modelizar considerando el sexo de cada progenitor, asumiendo la misma parametrización, aunque estimando una α para sementales (α_s) y otra para hembras reproductoras (α_h). A partir de un enfoque Bayesiano, las distribuciones posteriores conjuntas de los parámetros del modelo se definieron como:

$$p(\alpha|\mathbf{y}) \propto p(\mathbf{y}|\alpha)p(\alpha) \text{ y } p(\alpha_s, \alpha_h|\mathbf{y}) \propto p(\mathbf{y}|\alpha_s, \alpha_h)p(\alpha_s)p(\alpha_h),$$

donde \mathbf{y} era el vector de genotipos de la generación descendiente. Se asumieron a priori planos entre -0,5 y 0,5 para todos los parámetros de distorsión y una distribución multinomial para la verosimilitud de los datos (Casellas et al., 2014). Los parámetros α , α_s y α_h se caracterizaron a partir de sus distribuciones marginales posteriores obtenidas por Metropolis-Hastings (Hastings, 1970) con una única cadena de Markov Monte Carlo de 50.000 iteraciones. Su relevancia estadística se evaluó mediante un factor Bayes (FB).

RESULTADOS Y DISCUSIÓN

El FB compara la probabilidad del modelo con DS ($\alpha \neq 0$), en relación a otro modelo con $\alpha=0$. Podemos ver una relación directa del FB con la magnitud de DS en la Figura 2, donde a mayor DS más evidencia estadística se obtenía, decisiva a partir de $\text{FB} > 100$ ($\log_{10}(\text{FB}) > 2$) según la escala de Jeffreys (1984). Con el incremento de N también aumentó el FB (Figura 2), y con pocos individuos (N=25) los niveles bajos de DS no se pudieron detectar fácilmente (Tabla 1). Hay que mencionar que con escenarios de DS nula se obtuvo hasta un 0,29% de falsos positivos con $\text{FB} > 10$ y 0,02% con $\text{FB} > 100$. En general, los valores de FB de la α_s y α_h respecto la α fueron menos altos bajo la misma N (Figura 2, resultados no mostrados para

α_h). Esto último se vio influido por el número de parámetros que se estimaron con la misma cantidad de datos, donde el α dispone de dos fuentes de información (paterna y materna). La proporción de progenitores heterocigotos afectó claramente la potencia estadística del test. En la Figura 2 se presenta el FB en función de la proporción de progenitores heterocigotos (media entre la proporción de sementales y hembras heterocigotos; He) para DS y proporción de sementales heterocigotos para DS_s (He_s). A mayor He (o He_s) más evidencia estadística se observó con el FB. Los FB máximos se obtuvieron cuando la mayoría de los progenitores eran heterocigotos (He y $He_p > 0,5$), situación que se dio más fácilmente cuando habían pocos sementales en la población ($\text{♀}:\text{♂}=100$). En contra, cuando habían pocos sementales y todos eran homocigotos, la potencia disminuía tal como muestra la Figura 2F, y aún más para el caso de DS_s , siendo esta nula si no había ningún semental heterocigoto (Figura 2I; línea gris claro).

Por último, cabe destacar que la estimación de la DS mediante el modelo Bayesiano alcanzó una precisión (coeficiente de correlación entre los valores simulados y estimados de DS) superior a 0,97 para $FB > 10$. Sin embargo, la precisión de todas las réplicas sin descartar las no significativas fue más baja (0,60 en el caso de $N=100$ y $\text{♀}:\text{♂}=100$ para DS_s).

Podemos concluir que la potencia estadística del modelo Bayesiano de análisis aumenta con el tamaño de la población, la proporción de sementales heterocigóticos (para DS_s) y hembras heterocigóticas (para DS_h) o la media de los dos (para DS) en la generación parental y la magnitud de la distorsión de la segregación (Figura 1). Aunque el número de hembras por semental no afecta a la potencia estadística por sí mismo, sí influyó sobre el número de sementales, y en consecuencia sobre la variabilidad genética disponible y la proporción de sementales heterocigóticos de la población. El siguiente paso es validar los resultados obtenidos con datos reales en una población de vacuno de leche (Holstein) con 1.263.568 animales genotipados y fenotipados para caracteres relacionados con la fertilidad y reproducción en vacas y toros (incluyendo parámetros de calidad del semen) con el fin de identificar las regiones en el genoma con distorsión de la segregación vía macho y/o hembra en poblaciones con elevada potencia estadística.

REFERENCIAS BIBLIOGRÁFICAS

- Casellas, J. et al. 2014. *Genetics* 198: 1357-1367.
- Hastings, W.K. 1970. *Biometrika* 57: 97-109.
- Huang, L. O. et al. 2013. *Hum. Genet.* 132: 245–263.
- Jeffreys, H. 1984. Clarendon Press, Oxford, UK.
- Moore, C. 2006. *Mamm. Genome* 17: 1005–1012.
- Pardo-Manuel de Villena, F., et al. 2000. *Genetics* 154: 333–342.
- Silver, L. 1993. *Trends in Genetics* 9: 250–254.

Agradecimientos: Investigación encuadrada en el proyecto BFO-16-04 financiado por Beef Farmers Ontario, Semex and American Angus Association - Angus Genetics Inc.

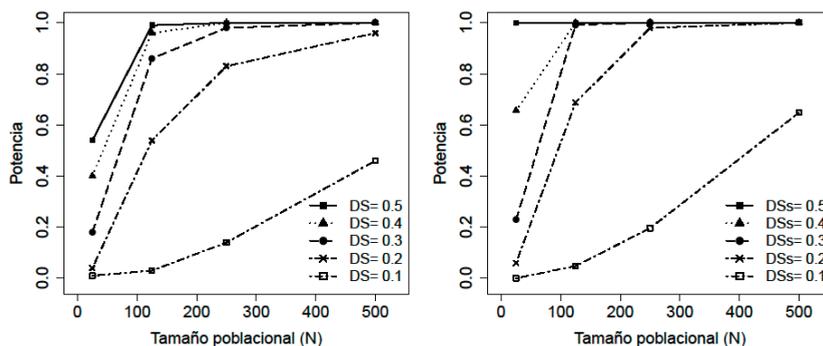


Figura 1. Potencia de detectar la distorsión de la segregación (DS; izquierda) o DS específica de sementales (todos heterocigotos; DS_s ; derecha) con un factor Bayes > 100 .

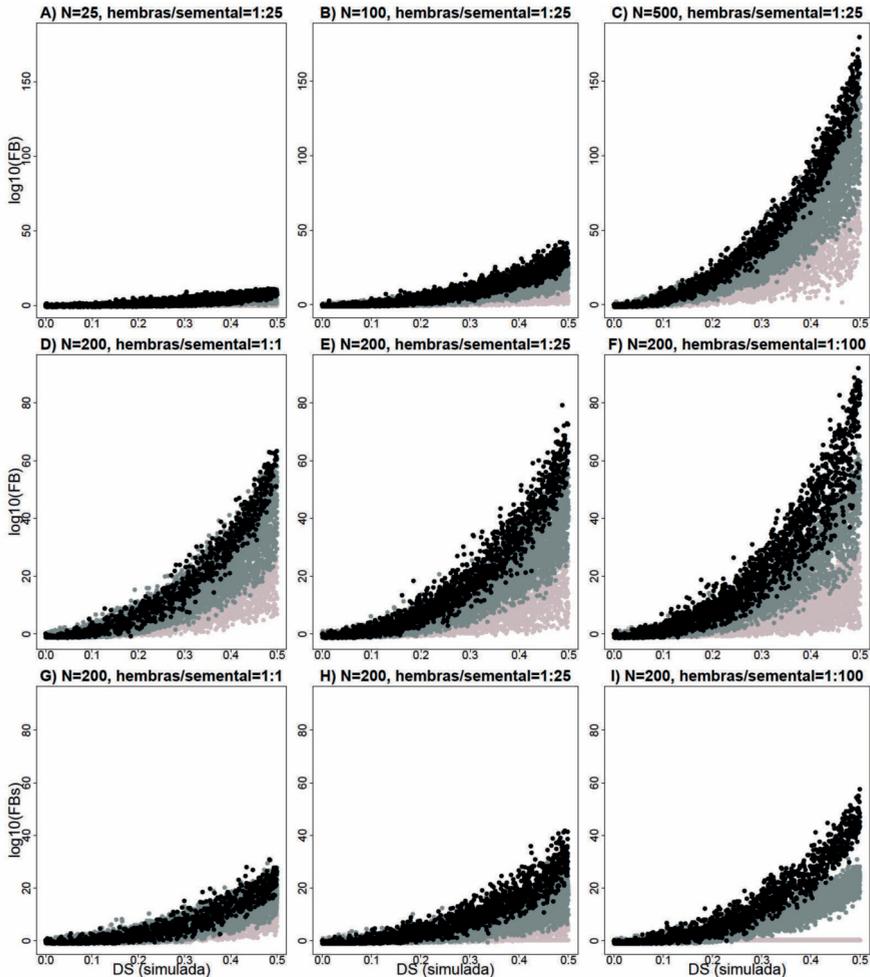


Figura 2. Factor Bayes (FB) en función de los valores absolutos de la distorsión de la segregación global (DS; A-F) o DS específica de sementales (DS_s ; G-I), y de la media entre la proporción intra-sexo de reproductores heterocigotos relevantes con valores $>0,5$ (negro), $\leq 0,5$ y $\geq 0,25$ (gris oscuro), o $<0,25$ (gris claro).

STATISTICAL POWER OF THE BAYESIAN ANALYSIS FOR TRANSMISSION RATIO DISTORTION

ABSTRACT: The aim of this research was to study the statistical power of the Bayesian analysis of transmission ratio distortion (TRD). The results obtained on simulated data sets showed that statistical power increased with population size, the mean between intra-sex proportion of relevant heterozygous parents, and the magnitude of the simulated TRD. The sire:dam ratio influenced the number of contributing sires and, consequently, the available genetic variability and the proportion of heterozygous sires. Pearson correlation coefficients between simulated and estimated TRD were greater than 0.97 when the Bayes Factor (BF) clearly favored the model with TRD against model without TRD ($BF > 10$).

Keywords: transmission ratio distortion, heterozygous progenitors, Bayes factor.