

## **DISTORSIÓN DE LA SEGREGACIÓN MENDELIANA EN EL GENOMA BOVINO: PATRONES DE NATURALEZA ALÉLICA Y GENOTÍPICA**

Id-Lahoucine S<sup>1.</sup>, Casellas, J<sup>2.</sup>, Suárez-Vega, A<sup>1.</sup>, Miglior, F<sup>1,3.</sup>, Fonseca, P<sup>1,4.</sup>, M. Sargolzaei M<sup>1,5.</sup>, Miller, S<sup>1,6.</sup>, Schenkel, F<sup>1.</sup>, Medrano, J<sup>7.</sup> y Cánovas, A<sup>1.</sup>

<sup>1</sup>Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, N1K 1E5 Guelph, Canada.

sidlahou@uoguelph.ca

<sup>2</sup>Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), España.

<sup>3</sup>Canadian Dairy Network, N1K 1E5 Guelph, Canada.

<sup>4</sup>Universidade Federal de Minas Gerais, Departamento de Biologia Geral, 31270-901 Belo Horizonte, Brazil.

<sup>5</sup>The Semex Alliance, N1G 3Z2 Guelph, Canada.

<sup>6</sup>Angus Genetics Inc., 64506 St. Joseph, MO, United States.

<sup>7</sup>University of California-Davis, Department of Animal Science, 95616 Davis, CA, USA.

### **INTRODUCCIÓN**

La distorsión de la segregación Mendeliana (DSM) es la desviación estadística de las frecuencias genotípicas esperadas bajo herencia Mendeliana (Silver, 1993). A nivel bilógico, la DSM se atribuye a mecanismos que afectan tanto a las células germinales (p. ej., deriva meiótica, selección de línea germinal, competencia gamética), como en el desarrollo embrionario o fetal (Huang et al., 2013), o incluso en la vida postnatal (Moore, 2006). Sin embargo, la DSM sigue siendo un fenómeno ambiguo (Nadeau, 2017) y las metodologías estadísticas específicas para sus análisis resultan limitadas y poco explotadas. Los modelos desarrollados por Casellas et al. (2012, 2014, 2017) proporcionan dos enfoques distintos para el estudio de la DSM. Mediante la parametrización alélica se puede trazar la transmisión de los alelos desde los progenitores heterocigotos a la progenie permitiendo así estimar una DSM global o diferenciarla vía paterna y materna. La parametrización genotípica modela la DSM basándose en los propios genotipos de la descendencia capturando interacciones entre los alelos y estimando un componente aditivo y dominante de la DSM. Los diferentes modelos se pueden ver como métodos analíticos complementarios que pueden contribuir con evidencias adicionales sobre la etapa reproductiva en que ocurre la DSM (p. ej., antes o después de la fecundación) y/o los mecanismos fisiológicos involucrados. En este contexto, el objetivo es este estudio era evaluar y comparar tres modelos de DSM sobre la raza Holstein.

### **MATERIAL Y MÉTODOS**

Los análisis incluyeron los genotipos de todos los tríos (toro-vaca-ternero) disponibles de la base de datos de Canadian Dairy Network (Guelph, ON) con descendencia genotipada dentro de los 3 primeros meses al nacimiento, evitando así animales pre-seleccionados. Se seleccionaron todos los marcadores con al menos 10 tríos, resultando en total 132.990 SNPs autosómicos distribuidos por todo el genoma (con un máximo de 340.363 tríos). El número de sementales varió entre 5 y 5.976 mientras el número de vacas oscilaba entre 7 y 132.282. El número total de animales genotipados usados fue 436.651.

Se implementaron 3 modelos para evaluar la DSM por cada marcador del genoma. El modelo genotípico con el componente aditivo y dominante de la DSM (Casellas et al., 2012), el modelo alélico con la DSM global (Casellas et al., 2014) y el modelo alélico con la DSM específica de sexo (paterna y materna) (Casellas et al., 2014, 2017). Los análisis se abordaron con un enfoque Bayesiano usando una única cadena Markov Monte Carlo de 110.000 iteraciones y siendo las 10.000 primeras iteraciones consideradas *burn-in*. La relevancia estadística de la DSM se evaluó mediante un factor de Bayes (BF; Kass y Raftery, 1995) y los tres modelos se compararon mediante el *deviance information criterion* (DIC; Spiegelhalter et al., 2002) para determinar la naturaleza de la DSM para cada marcador.

## RESULTADOS Y DISCUSIÓN

Los análisis de la DSM identificaron inicialmente 6.265 SNPs con al menos uno de los 3 modelos reportando evidencias estadísticas decisivas ( $BF \geq 100$ ) según la escala de Jeffreys (1984). Sin embargo, gran parte de estas regiones eran consecuencia del muestreo de los genomas (es decir, DSM aleatoria) o falso DSM. Hay que enfatizar que los errores de genotipado producen artefactos importantes en los análisis de la DSM, sobre todo cuando progenitores homocigotos se genotipan como heterocigotos. Dentro de este contexto, el número de marcadores se redujo a 5.618 una vez descartados los SNPs que presentaban un porcentaje de incompatibilidad superior al 5% y/o un limitando número de progenitores informativos ( $< 10$  sementales o/y vacas heterocigotos). Por otra parte, y para eliminar la DSM producida por azar, se redujo el número de marcadores siguiendo la distribución empírica de la DSM nula (Id-Lahoucine et al., 2018) a 3.277 y 547, permitiendo un margen de error del  $< 1\%$  y  $< 0.001\%$ , respectivamente.

La DSM tenían diferentes patrones genéticos. Comparando el DIC de los 3 modelos se pudo determinar la naturaleza de la DSM (global, específica de sexo o con componente aditivo y dominante) para cada región. En la tabla 1 se presenta una muestra de 10 regiones en base a la medida de ajuste (DIC) para los 3 modelos, así mismo, se muestra la magnitud de la DSM correspondiente al modelo con mejor ajuste, y la distribución de la progenie para cada tipo de cruce. El nivel de ajuste en las 10 regiones de la tabla 1 varía desde 0 hasta 1.305,6 unidades DIC. Por lo tanto, se puede observar que cuando no existe una DSM específica de sexo, la DSM global estimada debería ser idéntica a la DSM paternal y maternal (Id-Lahoucine et al., 2018), resultando en un ajuste prácticamente igual para ambos modelos. Fue por esto último que la mínima diferencia en unidades DIC se observó entre el modelo con DSM global y DSM específica de sexo, ante marcadores con DSM global. Sin embargo, cuando existía una DSM específica de sexo, la medida de ajuste entre ambos modelos resulto relevante reportando diferencias que oscilaban entre 45 y 76 unidades DIC y descartando así el modelo con DSM global (Tabla 1). Por otra parte, el modelo genotípico se vio favorecido en determinadas regiones genómicas, alcanzando ventajas en el ajuste de hasta 1.363 unidades DIC en relación al modelo alélico. La distribución de la progenie para los diferentes cruces de la tabla 1 muestra los distintos comportamientos de DSM por los correspondientes modelos.

En el genoma de Holstein se ha visto diferentes patrones de DSM. De las 547 regiones identificadas con DSM significativa, 199 SNPs se ajustaron al modelo alélico con DSM global. El modelo de DSM específico de sexo identificó 163 marcadores, siendo 57 de ellos específicos para sementales con una DSM materna nula. Del mismo modo, 44 SNPs mostraron una DSM materna y una DSM paterna nula. Las regiones restantes (62) presentaban tanto DSM paterna como materna, siendo parecidos a una DSM global, pero con diferentes magnitudes de DSM para cada progenitor. Finalmente, 185 regiones tenían DSM genotípica.

Los resultados de este estudio concluyen la necesidad de implementar los 3 modelos para capturar todo tipo de DSM y compararlos mediante métodos estadísticos apropiados para determinar la naturaleza de la DSM y así obtener información adicional para una mejor comprensión del fenómeno y los mecanismos biológicos subyacentes.

## REFERENCIAS BIBLIOGRÁFICAS

- Casellas, J. et al. 2012. *Genetics*. 191:247–259.
- Casellas, J. et al. 2014. *Genetics* 198: 1357-1367.
- Casellas, J. et al. 2017. *Anim. Genet.* 48:93-96.
- Huang, L. O. et al. 2013. *Hum. Genet.* 132: 245–263.
- Id-Lahoucine et al. 2018. (bajo revisión).
- Jeffreys, H. 1984. Clarendon Press, Oxford, UK.
- Kass, R. E. et al. 1995. *J. Am. Stat. Assoc.*
- Moore, C. 2006. *Mamm. Genome* 17: 1005–1012.
- Nadeau J. H. 2017. *Genetics*. 207:369-387.
- Silver, L. 1993. *Trends in Genetics* 9: 250–254.
- Spiegelhalter, D. J. Et al. 2002. *J. Royal Stat. Soc. B* 64: 583-639.

**Agradecimientos:** Esta investigación fue financiada por Beef Farmers Ontario (BFO-16-04) y Semex Alliance. Los autores reconocen los datos proporcionados por Canadian Dairy Network.

**Tabla 1.** Deviance information criterion (DIC) para cada modelo, distribución de la progenie para los diferentes cruces (semental x vaca) y la magnitud de la distorsión de la segregación (DSM) correspondiente al modelo ajustado para 10 regiones.

DIC M. Alélico DSM Global	DIC M. Alélico DSM de Progenitor	DIC M. DSM Genotípico	AB x AA*		AB x BB		AA x AB		BB x AB		AB x AB			Tipos y magnitudes de la DSM
			AA**	AB	AB	BB	AA	AB	AB	BB	AA	AB	BB	
<b>427,1</b>	427,7	459,6	2714	64	3	4	457	11	0	1	344	13	8	Global -0.47
<b>136,7</b>	137,5	148,1	0	0	20	879	0	0	7	346	1	5	64	Global 0.39
<b>354,4</b>	354,4	360,3	222	33	1	0	857	90	0	0	23	3	0	Global 0.24
1060,2	<b>910,4</b>	1060,9	1228	184	0	0	324	349	0	0	18	19	2	Pat. 0.37; mat. -0.01
2727,3	<b>2651,3</b>	2727,4	1	0	619	1534	27	42	1023	1074	46	117	84	Pat. -0.21; mat. -0.01
2169,7	<b>2124,4</b>	2170,2	4	4	725	749	6	3	603	1300	36	112	71	Pat. -0.01; mat. -0.18
1874,6	<b>1810,3</b>	1874,5	1097	884	0	0	864	260	0	0	55	54	11	Pat. 0.05; mat. 0.27
16884,1	16826,1	<b>15520,6</b>	3305	5122	161	77	3605	4778	949	299	1502	5724	510	Adi. -0.30; dom. 0.39
243,7	244,4	<b>236,8</b>	3	4	38	1088	0	0	8	289	6	8	79	Adi. 0.96; dom. -0.88
3902,9	3842,1	<b>2991,7</b>	14	92	1078	659	3	5	696	419	22	2505	487	Adi. 0.37; dom. 0.57

\*Genotipo semental x vaca. \*\*genotipo progenie. M.: modelo, pat.: paternal, mat.: maternal, adi.: aditivo, dom.: dominante.

#### TRANSMISSION RATIO DISTORTION ACROSS CATTLE GENOME: PATTERNS OF ALLELIC AND GENOTYPIC NATURE

**ABSTRACT:** Transmission ratio distortion (TRD) is attributed to various biological mechanisms acting on germ cells (e. g., meiotic drive, germline selection, gametic competition), at embryonic or fetal development or in the postnatal life. Three models of TRD based on allelic or genotypic parametrization were implemented and compared by deviance information criterion (DIC). SNP-by-SNP analyses were performed across the whole genome, including 436,651 genotypes of Holstein cattle. Results showed variation in the goodness-of-fit of the models across regions reporting differences up to 1,305.6 units of DIC. Across 132,282 autosomal SNPs, 199 markers were adjusted with the allelic model with overall TRD, 163 regions with parent-specific TRD and 185 SNPs with additive and dominance components of TRD. The results of this study conclude the importance of implementing different parameterization to capture all kinds of TRD and to compare those using appropriate statistical methods to determine the nature of the regions according to the goodness-of-fit.

**Keywords:** transmission ratio distortion, parent-specific, additive and dominance components, deviance information criterion.