



Definición y análisis del panel de polimorfismos de nucleótido simple a utilizar en pruebas de paternidad para tres razas de bovinos



Joel Domínguez-Viveros ^{a*}

Adán Medellín-Cazares ^a

Nelson Aguilar-Palma ^a

Francisco Joel Jahuey-Martínez ^a

Felipe Alonso Rodríguez-Almeida ^a

^a Universidad Autónoma de Chihuahua. Facultad de Zootecnia y Ecología. Periférico Francisco R. Almada km 1. 31453, Chihuahua, Chih. México.

*Autor de correspondencia: joeldguezviveros@yahoo.com.mx; jodominguez@uach.mx

Resumen:

Con el objetivo de definir el panel de polimorfismos de nucleótido simple (SNP) para pruebas de paternidad en bovinos, se analizaron los genotipos en tres razas (número de SNP evaluados e individuos muestreados): Hereford (HER; 202; 1317), Brangus (BRA; 217; 3431) y Limousin (LIM; 151; 8205). Dentro de raza, se descartó los SNP con porcentaje de individuos genotipados (PIG) menor a 90 %, con desequilibrio Hardy Weinberg (HW; $P < 0.05$), con frecuencia de alelo menor de 0.10 o menos y con desequilibrio de ligamiento, donde la correlación entre frecuencias genotípicas fue superior a 0.25. Se estimó los niveles de heterocigosis esperada (H_e) y observada (H_o), contenido de información polimórfica (CIP); así como, el índice de Shannon, el índice de fijación y tamaño efectivo de población (N_e). Se calculó la probabilidad de exclusión (PEC) y de identidad combinada (PIC). El panel final fue de 121, 188 y 113 SNP en HER, BRA y LIM, respectivamente; la principal fuente de descarte fue HW seguido de PIG. Los niveles de H_o y H_e fueron superior a 0.40; el PIC fue mayor a 0.32 y N_e presentó estimaciones por arriba de 181.3. Los resultados para la PEC fueron superiores a 0.999999; para la PIC, estuvieron por debajo de 1×10^{-20} .

Palabras clave: Heterocigosis, Probabilidad de exclusión, Probabilidad de identidad, Polimorfismo, Índice Shannon.

Recibido: 17/08/2020

Aceptado: 03/12/2020

En México, las evaluaciones genéticas (EVGE) en bovinos para carne se realizan a partir de 2001; en torno a 25 razas, dispuestas en asociaciones nacionales de criadores de ganado registro, las EVGE conjugan la información genealógica y productiva contenida en los libros de registro⁽¹⁾. La información genealógica, que conforma el registro genealógico de pureza de raza o grados de pureza, define las relaciones de parentesco de toda la población a través del pedigrí de cada individuo. Errores en la veracidad e integridad del pedigrí tiene efectos en la certeza de la pureza racial; en la definición de ancestros fundadores y asignación de individuos a generaciones, así como en los cálculos de los niveles de consanguinidad y parentesco^(2,3,4). En las EVGE, errores en la información genealógica tiene consecuencias en la estimación de componentes de varianza y parámetros genéticos, así como en la predicción de valores genéticos y jerarquización de sementales; por consecuencia, también afecta la respuesta a la selección y progreso genético^(5,6,7,8).

Los marcadores genéticos (MAGE) expresan el polimorfismo del ADN, su evolución y uso han fortalecido los programas de mejoramiento genético animal^(9,10,11). En bovinos, las pruebas de paternidad han evolucionado con el desarrollo de los MAGE⁽¹²⁾; la Sociedad Internacional de Genética Animal (SIGA) inicialmente propuso un panel de 121 SNP (Polimorfismo de Nucleótido Simple) desarrollado en razas *Bos taurus*, posteriormente se agregaron 100 SNP derivados de razas *Bos indicus*^(13,14). En México, las pruebas de paternidad se han implementado en las razas Brangus, Limousin y Hereford con base en el panel de SNP propuesto por la SIGA; no obstante, se requiere validar el panel de SNP a través de poblaciones, dado que la funcionalidad y veracidad de un MAGE en pruebas genéticas depende del equilibrio Hardy-Weinberg, el posible desequilibrio de ligamiento, el contenido de información polimórfica, entre otros componentes; además, en un conjunto de MAGE el poder de prueba es validado por la probabilidad de exclusión^(14,15).

Al respecto, se han realizado estudios validando el panel de SNP desarrollado por la SIGA a utilizar en pruebas de paternidad de bovinos en Brasil⁽¹⁶⁾, Argentina⁽¹⁷⁾, China^(18,19), Estados Unidos^(20,21), Japón^(22,23) y Europa^(24,25,26). Con base en lo anterior, los objetivos del presente estudio fueron validar el panel de SNP definido por la Sociedad Internacional de Genética Animal para pruebas genéticas en poblaciones de bovinos mexicanas.

Se analizaron los genotipos de SNP para bovinos: Brangus (BRA), Hereford (HER) y Limousin (LIM); en el Cuadro 1 se describe la base de datos analizada. En una primera edición se realizó un control de calidad de la base de datos; se verificó la información del individuo y de la muestra, así como conflicto mendeliano, genotipos duplicados e idénticos por estado. El panel evaluado en cada raza es un subconjunto del panel general propuesto por la SIGA; para LIM, el procesamiento de las muestras lo realizó el laboratorio Labogena con base en los SNP utilizados en Francia; para las otras razas, el proceso lo realizó el laboratorio Neogen GeneSeek con el conjunto de SNP utilizados en EU. Los análisis se desarrollaron dentro de raza en cuatro etapas.

1. Valoración del porcentaje de individuos (*call rate*) con genotipo identificado (PIG); estimación de las frecuencias alélicas y genotípicas, así como el análisis de equilibrio Hardy Weinberg (HW).
2. Descartando los SNP con desequilibrio HW ($P < 0.05$) y PIG menor a 90 %, se analizó el posible desequilibrio de ligamiento (DL) con base en la correlación (r^2) entre frecuencias genotípicas a través de SNP, se estimó la heterocigosis esperada (H_e) y observada (H_o), el contenido de información polimórfica (CIP), el índice de Shannon (IS) y el índice de fijación (FIS). Con la r^2 promedio y ajustada por el tamaño de muestra se estimó el tamaño efectivo de la población (N_e), con base en el planteamiento de Waples⁽²⁷⁾.
3. Se integró un panel de SNP por raza descartando los SNP con desequilibrio HW ($p < 0.05$), con frecuencia de alelo menor (FAM) igual o menor a 0.10, con DL⁽²⁶⁾ donde la r^2 fue superior a 0.25 y PIG menor a 0.90 %.
4. Con el subconjunto de SNP para cada raza, se ordenaron de forma descendente por CIP y se calculó la probabilidad de exclusión (PE) en tres modalidades^(28,29,30): a) con un progenitor candidato y otro conocido, para excluir al progenitor candidato [$PE1 = 1 - 2 * \sum_{i=1}^n p_i^2 + \sum_{i=1}^n p_i^3 + 2 * \sum_{i=1}^n p_i^4 - 3 * \sum_{i=1}^n p_i^5 - 2 * (\sum_{i=1}^n p_i^2)^2 + 3 * \sum_{i=1}^n p_i^2 * \sum_{i=1}^n p_i^3$]; b) dado un progenitor candidato y la progeñie, poder excluir la relación entre ellos [$PE2 = 1 - 4 * \sum_{i=1}^n p_i^2 + 2 * (\sum_{i=1}^n p_i^2)^2 + 4 * \sum_{i=1}^n p_i^3 - 3 * \sum_{i=1}^n p_i^4$]; y, c) con dos progenitores candidatos, exclusión de uno o ambos [$PE3 = 1 + 4 * \sum_{i=1}^n p_i^4 - 4 * \sum_{i=1}^n p_i^5 - 3 * \sum_{i=1}^n p_i^6 - 8 * (\sum_{i=1}^n p_i^2)^2 + 8 * (\sum_{i=1}^n p_i^2) * (\sum_{i=1}^n p_i^3) + 2 * (\sum_{i=1}^n p_i^3)^2$]. La probabilidad de exclusión combinada para cada situación fue ($PEC = 1 - \Pi(1 - PE_i)$). Además, se estimaron dos probabilidades de identidad (PI)⁽³¹⁾: la probabilidad de identidad de dos individuos tomados al azar, presenten genotipos idénticos [$PI1 = \sum_{i=1}^n p_i^4 + \sum_{i=1}^n \sum_{j=1}^n (2p_i p_j)^2$]; y, la probabilidad de identidad para dos hermanos completos, tomados al azar, presenten genotipos idénticos [$PI2 = 0.25 + (0.5 * \sum_{i=1}^n p_i^2) + (0.5 * (\sum_{i=1}^n p_i^2)^2) - (0.25 * \sum_{i=1}^n p_i^4)$]. La probabilidad de identidad combinada (PIC) para cada situación, se calculó con el producto de las probabilidades de identidad de cada marcador. Los análisis se realizaron con los programas FSTAT⁽³²⁾, LDNE⁽³³⁾ y GenAlex⁽³⁴⁾.

En el Cuadro 1 se resume el proceso de selección y descarte de SNP a través de raza, así como la estructura del panel final. El total de SNP retirados por raza, como porcentaje del total evaluados, fluctuó de 13.4 % (BRA) a 40.0 % (HER), donde la principal causa de descarte fue el desequilibrio HW ($P < 0.05$). En el proceso de descarte de SNP no se observó una tendencia o asociación entre marcadores, el conjunto de SNP separado a través de razas fue diferente. El número final de SNP por raza fluctuó de 113 (LIM) a 188 (BRA), los cuales están dentro de los lineamientos de la SIGA⁽¹³⁾, la cual estipula que el panel por raza debe estar conformado por al menos 100 SNP.

Cuadro 1: Definición del panel de SNP por raza con base en los criterios de descarte

Raza	N	SNPn	PIG	HW	FAM	DL	SNPf
Herford	1,317	202	41	30	8	2	121
Brangus	3,431	217	2	19	2	6	188
Limousin	8,205	151	9	28	1	0	113

N= número de individuos muestreados. SNPn= número de SNP evaluados. PIG= número de SNP retirados por porcentaje de individuos con genotipos identificados menor a 90%. HW= número de SNP descartados por presentar desequilibrio Hardy Weinberg ($P < 0.05$). FAM= número de SNP separados por frecuencia de alelo menor, menor a 0.10. DL= número de SNP descartados por desequilibrio de ligamiento, dado que la correlación entre frecuencias fue superior a 0.25. SNPf= total de SNP que conforman el panel por raza.

En el Cuadro 2, se presentan los resultados para H_o , H_e , CIP, FIS y N_e . No se observan diferencias entre H_o y H_e , lo cual refleja que el conjunto de SNP seleccionado está en equilibrio HW. Para el IS, los resultados en las tres poblaciones fueron por debajo de la unidad, lo cual se puede asociar con homogeneidad en las poblaciones y se reduce la incertidumbre para predecir la probabilidad de asignación de un individuo a la población que pertenecerá. Para el FIS, todos los resultados tienden a cero, señalando una estabilidad en la relación de homocigotos y heterocigotos. Con las estimaciones del N_e , en el marco del equilibrio HW los incrementos esperados de consanguinidad ($\Delta F = 1 / 2N_e$) por generación van del 0.08 al 0.27 %. Los niveles de H_e , H_o y CIP determinan si un marcador genético es o no informativo y su potencial de uso en estudios de variabilidad genética; no obstante, la jerarquización u ordenamiento de los SNP por la capacidad de uso puede ser diferente a través de poblaciones.

Cuadro 2: Indicadores de variabilidad genética (valores promedio) con base en el panel de SNP seleccionado para cada raza

Raza	H_o	H_e	CIP	IS	FIS	N_e
Herford	0.416	0.419	0.328	0.607	0.008	181.3
Brangus	0.433	0.434	0.337	0.623	0.002	246.9
Limousin	0.451	0.452	0.348	0.643	0.004	629.8

Heterocigosis observada (H_o) y esperada (H_e). CIP= contenido de información polimórfica. IS= Índice Shannon. FIS= índice de fijación. N_e = tamaño efectivo.

Con el total de SNP seleccionado en cada raza, los resultados para la PEC en las tres modalidades fueron superiores a 0.999999; para la PIC, estuvieron por debajo de 1×10^{-39} y 1×10^{-20} en PI1 y PI2, respectivamente. En el Cuadro 3 se describen los resultados para las formas alternas de PEC y PIC, parcialmente alcanzados con 50 SNP. Dado la estructura genética de las poblaciones y las fuerzas que inciden sobre la genética de poblaciones, la conformación y arreglo de un panel de SNP para verificar paternidad en bovinos puede tener diversas dimensiones y valores de probabilidad: Heaton *et al*⁽²⁰⁾ con un panel de 32 SNP a través de 17 razas, publicaron PEC superior a 0.994 y PIC de 1.9×10^{-13} ; Van Eenennaam *et al*⁽²¹⁾ con 28 SNP, FAM superior a 0.40 en hatos comerciales, obtuvieron PEC de 0.956; Hara *et al*⁽²⁹⁾ con 29 SNP para una raza nativa de Japón reportaron PIC de 2.73×10^{-12} y PEC de 0.96929 a 0.99693. En otros estudios afines, Werner *et al*⁽²⁴⁾ publicaron PEC superior a 0.9999 y PIC de 1×10^{-13} con 37 SNP. Fernández *et al*⁽¹⁷⁾ en Angus con un arreglo de 116 SNP, reportaron probabilidades de no exclusión combinada ($PNEC = 1 - PEC$) en el intervalo de 2.1×10^{-4} a 1.4×10^{-9} , así como PIC de 4.1×10^{-15} . Panetto *et al*⁽¹⁶⁾ para la raza Sindhi de Brasil, con 71 SNP donde la FAM fue superior a 0.35, publicaron PNEC de 1×10^{-8} . Zhang *et al*⁽¹⁸⁾ en ganado Simmental con 50 SNP y FAM superior a 0.40, reportaron PEC superior a 0.9989; Hu *et al*⁽¹⁹⁾ en bovinos cruzados de China, con 50 SNP donde el valor promedio de la FAM fue de 0.43, obtuvieron PEC de 0.99797 a 0.999999.

Cuadro 3: Valores de probabilidad de exclusión y de identidad, obtenidos con 50 SNP dentro del panel total seleccionado por raza

Raza	PEC1	PEC2	PEC3	SNPi	PIC1	PIC2
Hereford	0.99996	0.99831	0.99999	113	1.0E-21	8.2E-12
Brangus	0.99996	0.99861	0.99999	91	6.2E-22	5.7E-12
Limousin	0.99996	0.99849	0.99999	97	7.7E-22	6.6E-12

PEC1= probabilidad de exclusión combinada, con un progenitor candidato y otro conocido. PEC2= probabilidad de exclusión combinada, dado un progenitor candidato y la progeñe. PEC3= probabilidad de exclusión combinada con dos progenitores candidatos. PIC1= probabilidad de identidad combinada para dos individuos tomados al azar. PIC2= probabilidad de identidad combinada, para dos hermanos completos tomados al azar. SNPi= número de SNP requeridos para obtener un valor superior a 0.999999 en las probabilidades de exclusión.

Para bovinos Brangus, Hereford y Limousin el número de SNP que conforma el panel para pruebas de paternidad fue superior a 100; seleccionados con base en los criterios asociados a variabilidad genética y estructura de la población, con valores de probabilidad de exclusión superior a 0.999999 y probabilidad de identidad por debajo a 6.6×10^{-12} .

Agradecimientos

Se agradece a la asociación Nacional de Criadores de Bovinos Brangus y Limousin; así como a la Hereford Mexicana, por proporcionar la base de datos del presente estudio. Se agradece al Consejo Nacional de Ciencia y Tecnología por la beca para estudios de posgrado proporcionada.

Todos los autores declaran que no existe conflicto de interés alguno.

Literatura citada:

1. CONARGEN. Guía técnica de programas de control de producción y mejoramiento genético en bovinos. Consejo Nacional de los Recursos Genéticos Pecuarios. México. 2010.
2. Banos G, Wiggans GR, Powell RL. Impact of paternity errors in cow's identification on genetic evaluations and international comparisons. *J Dairy Sci* 2001;84:2523-2529.
3. Atkin FC, Dieters MJ, Stringer JK. Impact of depth of pedigree and inclusion of historical data on the estimation of additive variance and breeding values in a sugarcane breeding program. *Theo Appl Gen* 2009;119:555-565.
4. Ramírez-Valverde R, Delgadillo-Zapata AR, Domínguez-Viveros J, Hidalgo-Moreno JA, Núñez-Domínguez R, Rodríguez-Almeida FA, *et al.* Análisis del pedigrí en la determinación de la diversidad genética de poblaciones bovinas para carne mexicana. *Rev Mex Cienc Pecu* 2018;9:614-635.
5. Visscher PM, Woolliams JA, Smith D, Williams JL. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. *J Dairy Sci* 2002;85:2368-2375.
6. Sanders K, Bennewitz J, Kalm E. Wrong and missing sire information affects genetic gain in the Angeln dairy cattle population. *J Dairy Sci* 2006;89:315-321.
7. Parlato E, Van Vleck LD. Effect of parentage misidentification on estimates of genetic parameters for milk yield in the Mediterranean Italian buffalo population. *J Dairy Sci* 2012;95:4059-4064.
8. Raoul J, Palhiere I, Astruc JM, Elsen JM. Genetic and economic effects of the increase in female paternal filiations by parentage assignment in sheep and goat breeding programs. *J Anim Sci* 2016;94:3663-3683.

9. Vignal A, Milan D, SanCristobal M, Eggen A. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol* 2002;34:275-305.
10. Cañón J. Using molecular information in animal improvement programs. *Rev Corpoica* 2006;7:5-15.
11. Dekkers, JCM. Application of genomics tools to animal breeding. *Current Genomics* 2012;13:207-212.
12. Flanagan SP, Jones AG. The future of parentage analysis: from microsatellites to SNPs and beyond. *Mol Ecol* 2019;28:544-567.
13. Morrin R, Boscher M. Cattle molecular markers and parentage testing workshop. ISAG Conference 2012;1-7.
14. Strucken EM, Lee SH, Lee HK, Song KD, Gibson JP, Gondro C. How many markers are enough? Factors influencing parentage testing in different livestock populations. *J Anim Breed Genet* 2016;133:13-23.
15. Baruch E, Weller J. Estimation of the number of SNP genetic markers required for parentage verification. *Anim Genet* 2008;39:474-479.
16. Panetto JCD, Machado MA, da Silva MVG, Barbosa RS, dos Santos GG, Leite RMHR, Peixoto MGC. Parentage assignment using SNP markers, inbreeding and population size for the Brazilian Red Sindhi cattle. *Livest Sci* 2017;204:33-38.
17. Fernández ME, Goszczynski DE, Liron JP, Villegas-Castagnasso EE, Cariño MH, Ripoli MV, *et al.* Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability, and assessment parentage in an inbred Angus herd. *Genet Mol Biol* 2013;36:185-191.
18. Zhang T, Guo L, Shi M, Xu L, Chen Y, Zhang L, Gao H, Li J, Gao X. Selection and effectiveness of informative SNPs for paternity in Chinese Simmental cattle based on a high-density SNP array. *Gene* 2018;673:211-216.
19. Hu L, Li D, Chu Q, Wang Y, Zhou L, Yu Y, Zhang Y, *et al.* Selection and implementation of SNP markers for parentage analysis in a Chinese crossbred cattle population. *Res Square* 2020;e30446/v1.
20. Heaton MP, Harhay GP, Bennett GL, Stone RT, Grosse WM, Casas E, *et al.* Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. *Mamm Gen* 2002;13:272-281.

21. Van Eenennaam AL, Weaber RL, Draker DJ, Penedo MCT, Quaas RL, Garrick DJ, Pollak EJ. DNA-based paternity analysis and genetic evaluation in a large commercial cattle ranch setting. *J Anim Sci* 2007;85:3159-3169.
22. Honda T, Katsuta T, Mukai F. Simulation study on parentage analysis with SNPs in the Japanese cattle population. *Asian-Aust J Anim Sci* 2009;10:1351-1358.
23. Strucken EM, Gudex B, Ferdosi MH, Lee HK, Song KD, Gibson JP, *et al.* Performance of different SNP panels for parentage testing in two East Asian cattle breeds. *Anim Genet* 2014;45:572-575.
24. Werner FAO, Durstewitz G, Habemann FA, Thaller G, Kramer W, Kollers S, *et al.* Detection, and characterization of SNP useful for identity control and parentage testing in major European dairy breeds. *Anim Genet* 2004;35:44-49.
25. Negrini R, Nicoloso L, Crepaldi P, Milanesi E, Colli L, Chegdani F, *et al.* Assessing SNP markers for assigning individuals to cattle populations. *Anim Genet* 2009;40:18-26.
26. Allen AR, Taylor M, McKeown B, Curry AI, Lavery JF, Mitchell A, *et al.* Compilation of a panel of informative single nucleotide polymorphisms for bovine identification in the Northern Irish cattle population. *BMC Genet* 2010;11:Art 5.
27. Waples RS. A bias correction for estimate of effective population size base on linkage disequilibrium at unlinked loci. *Conserv Genet* 2006;7:167-184.
28. Jamieson A, Taylor SC. Comparisons of three probability formulae for parentage exclusion. *Anim Genet* 1997;28:397-400.
29. Hara K, Watabe H, Sasazaki S, Mukai F, Mannen H. Development of SNP markers for individual identification and parentage test in Japanese black cattle population. *Anim Sci J* 2010;81:152-157.
30. Olenski K, Kaminski S, Tokarska M, Hering DM. Subset of SNPs for parental identification in European bison Lowland-Bialowieza line (*Bison bonasus bonasus*). *Conserv Genet Res* 2018;10:73-78.
31. Waits L, Luikart G, Taberlet P. Estimating the probability of identity among genotypes in natural populations cautions and guidelines. *Mol Ecol* 2001;10:249-256.
32. Goudet J. FSTAT: A computer program to calculate F-Statistics. *J Heredity* 1995;86:485-486.

33. Waples RS, Do Chi. LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Mol Ecol* 2008;8:753-756.
34. Peakall R, Smouse PE. GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* 2012;28:2537-2539.