https://doi.org/10.22319/rmcp.v12i3.5771

Technical note

Definition and analysis of the panel of SNPs to be used in paternity tests for three breeds of cattle

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.

*Corresponding author: joeldguezviveros@yahoo.com.mx; jodominguez@uach.mx

Abstract:

In order to define the SNP panel for paternity tests in cattle, genotypes were analyzed in three breeds (number of SNPs evaluated and individuals sampled): Hereford (HER; 202; 1317), Brangus (BRA; 217; 3431) and Limousin (LIM; 151; 8205). Within breed, SNPs with a percentage of genotyped individuals (PGI) less than 90 %, with Hardy-Weinberg disequilibrium (HW; $P<0.05$), with allele frequency less than 0.10 or less and with linkage disequilibrium, where the correlation between genotypic frequencies was greater than 0.25, were discarded. The levels of expected (He) and observed (Ho) heterozygosity, polymorphic information content (PIC) were estimated; as well as the Shannon index, the fixation index and effective population size (Ne). The combined exclusion probability (CEP) and identity probability (CIP) were calculated. The final panel was 121, 188 and 113 SNPs in HER, BRA and LIM, respectively; the main source of discard was HW followed by PGI. Levels of Ho and He were above 0.40; CIP was greater than 0.32 and Ne presented estimates above 181.3. The results for CEP were higher than 0.999999; for CIP, they were below 1×10^{-20}.

Key words: Heterozygosity, Exclusion probability, Identity probability, Polymorphism, Shannon Index.

Received: 17/08/2020

Accepted: 03/12/2020

In Mexico, genetic evaluations (GEEV) in beef cattle have been carried out since 2001; around 25 breeds, arranged in national associations of registered cattle breeders, GEEVs combine the genealogical and productive information contained in the registration books[(1)]. Genealogical information, which makes up the genealogical record of breed purity or degrees of purity, defines the parentage relationships of the entire population through the pedigree of each individual. Errors in the veracity and integrity of the pedigree have effects on the certainty of breed purity; in the definition of founding ancestors and assignment of individuals to generations, as well as in the calculations of the levels of consanguinity and parentage[(2,3,4)]. In GEEV, errors in genealogical information have consequences in the estimation of variance components and genetic parameters, as well as in the prediction of genetic values and hierarchization of sires; consequently, they also affect the response to genetic selection and progress[(5,6,7,8)].

Genetic markers (GEMA) express the polymorphism of DNA, their evolution and use have strengthened animal genetic improvement programs[(9,10,11)]. In cattle, paternity tests have evolved with the development of GEMAs[(12)]; the International Society for Animal Genetics (ISAG) initially proposed a panel of 121 SNPs (Single-Nucleotide Polymorphism) developed in *Bos taurus* breeds, later, 100 SNPs derived from *Bos indicus* breeds[(13,14)] were added. In Mexico, paternity tests have been implemented in the Brangus, Limousin and Hereford breeds based on the SNP panel proposed by ISAG; however, it is necessary to validate the SNP panel by populations, since the functionality and veracity of a GEMA in genetic tests depends on the Hardy-Weinberg equilibrium, the possible linkage disequilibrium, the polymorphic information content, among other components; in addition, in a set of GEMA, the test power is validated by the exclusion probability[(14,15)].

In this regard, studies have been carried out validating the SNP panel developed by ISAG to be used in cattle paternity tests in Brazil[(16)], Argentina[(17)], China[(18,19)], the United States[(20,21)], Japan[(22.23)] and Europe[(24,25,26)]. Based on the above, the objectives of this study were to validate the SNP panel defined by the International Society for Animal Genetics for genetic tests in Mexican cattle populations.

The genotypes of SNP for cattle were analyzed: Brangus (BRA), Hereford (HER) and Limousin (LIM); Table 1 describes the database analyzed. In a first edition, a quality control of the database was carried out; the information of the individual and the sample was verified, as well as Mendelian conflict, duplicate and identical genotypes by state. The panel evaluated in each breed is a subset of the general panel proposed by ISAG; for LIM, the processing of the samples was carried out by the Labogena laboratory based on the SNPs used in France; for the other breeds, the process was carried out by the Neogen GeneSeek laboratory with the set of SNPs used in the US. The analyses were developed within breed in four stages:

1. Assessment of the percentage of individuals (call rate) with identified genotype (PGI); estimation of allelic and genotypic frequencies, as well as Hardy-Weinberg (HW) equilibrium analysis.
2. Discarding the SNPs with HW disequilibrium ($P< 0.05$) and PGI less than 90 %, the possible linkage disequilibrium (LD) was analyzed based on the correlation (r^2) between genotypic frequencies through SNP, the expected (He) and observed heterozygosity (Ho), the polymorphic information content (PIC), the Shannon index (SI) and the fixation index (FIS) were estimated. With the average r^2 and adjusted for the sample size, the effective population size (Ne) was estimated, based on the Waples approach[27].
3. A panel of SNPs by breed was integrated, discarding SNPs with HW disequilibrium ($p< 0.05$), with lesser allele frequency (LAF) equal to or less than 0.10, with LD[26] where r^2 was greater than 0.25 and PGI less than 0.90 %.
4. With the subset of SNPs for each breed, they were sorted in descending order by PIC and the exclusion probability (EP) was calculated in three modalities[28,29,30]: (a) with one candidate parent and another known parent, to exclude the candidate parent [$EP1 = 1 - 2*\sum_{i=1}^{n} pi^2 + \sum_{i=1}^{n} pi^3 + 2*\sum_{i=1}^{n} pi^4 - 3*\sum_{i=1}^{n} pi^5 - 2*(\sum_{i=1}^{n} pi^2)^2 + 3*\sum_{i=1}^{n} pi^2*\sum_{i=1}^{n} pi^3$]; (b) given a candidate parent and the progeny, to be able to exclude the relationship between them [$EP2 = 1 - 4*\sum_{i=1}^{n} pi^2 + 2*(\sum_{i=1}^{n} pi^2)^2 + 4*\sum_{i=1}^{n} pi^3 - 3*\sum_{i=1}^{n} pi^4$]; and, (c) with two candidate parents, exclusion of one or both [$EP3 = 1 + 4*\sum_{i=1}^{n} pi^4 - 4\sum_{i=1}^{n} pi^5 - 3*\sum_{i=1}^{n} pi^6 - 8*(\sum_{i=1}^{n} pi^2)^2 + 8*(\sum_{i=1}^{n} pi^2)*(\sum_{i=1}^{n} pi^3) + 2*(\sum_{i=1}^{n} pi^3)^2$. The combined exclusion probability for each situation was ($CEP= 1 - \Pi(1 - EP_i)$. In addition, two identity probabilities (IP) were estimated[31]: the probability of identity of two individuals taken at random, present identical genotypes [$IP1 = \sum_{i=1}^{n} pi^4 + \sum_{i=1}^{n}\sum_{i=1}^{n}(2pipj)^2$]; and, the probability of identity for two full siblings, taken at random, present identical genotypes [$IP2 = 0.25 + (0.5*\sum_{i=1}^{n} pi^2) + (0.5*(\sum_{i=1}^{n} pi^2)^2) - (0.25*\sum_{i=1}^{n} pi^4)$]. The combined identity probability (CIP) for each situation was calculated with the product of the probabilities of identity of each marker. The analyses were performed with the programs FSTAT[32], LDNE[33] and GenAlex[34].

Table 1 summarizes the process of selecting and discarding SNPs by breed, as well as the structure of the final panel. The total number of SNPs removed by breed, as a percentage of

the total evaluated, fluctuated from 13.4 % (BRA) to 40.0 % (HER), where the main cause of discarding was the HW disequilibrium ($P<0.05$). In the process of discarding SNPs, no trend or association between markers was observed, the set of SNPs separated by breeds was different. The final number of SNPs per breed fluctuated from 113 (LIM) to 188 (BRA), which are within the guidelines of ISAG[13], which stipulates that the panel per breed must be made up of at least 100 SNPs.

Table 1: Definition of the SNP panel by breed based on discard criteria

Breed	N	SNPn	PGI	HW	LAF	LD	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= number of individuals sampled. SNPn= number of SNPs evaluated. PGI= number of SNPs removed due to percentage of individuals with identified genotypes less than 90 %. HW= number of SNPs discarded for presenting Hardy-Weinberg disequilibrium ($P<0.05$). LAF= number of SNPs separated due to lesser allele frequency, less than 0.10. LD = number of SNPs discarded due to linkage disequilibrium, since the correlation between frequencies was greater than 0.25. SNPf= total SNPs that make up the panel by breed.

Table 2 presents the results for Ho, He, PIC, FIS and Ne. No differences between Ho and He are observed, which reflects that the selected SNP set is in HW equilibrium. For SI, the results in all three populations were below one, which can be associated with homogeneity in the populations and the uncertainty to predict the probability of assigning an individual to the population that will belong reduces. For FIS, all results tend to zero, indicating a stability in the relationship of homozygotes and heterozygotes. With Ne estimates, within the framework of the HW equilibrium, the expected increases in consanguinity ($\Delta F = 1 / 2Ne$) per generation range from 0.08 to 0.27 %. He, Ho, and PIC levels determine whether or not a genetic marker is informative and its potential for use in genetic variability studies; however, the hierarchization or ordering of SNPs by the capacity of use may be different between populations.

Table 2: Indicators of genetic variability (average values) based on the SNP panel selected for each breed

Breed	Ho	He	PIC	SI	FIS	Ne
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

Observed (Ho) and expected (He) heterozygosity. PIC= polymorphic information content. SI= Shannon Index. FIS= fixation index. Ne= effective size.

With the total number of SNPs selected in each breed, the results for CEP in the three modalities were greater than 0.999999; for CIP, they were below 1×10^{-39} and 1×10^{-20} in

PI1 and PI2, respectively. Table 3 describes the results for the alternate forms of CEP and CIP, partially achieved with 50 SNPs. Given the genetic structure of the populations and the forces that affect the genetics of populations, the conformation and arrangement of a panel of SNPs to verify paternity in cattle can have different dimensions and probability values: Heaton *et al*[20], with a panel of 32 SNPs by 17 breeds, published a CEP greater than 0.994 and a CIP of 1.9 x 10^{-13}; Van Eenennaam *et al*[21], with 28 SNPs, LAF greater than 0.40 in commercial herds, obtained a CEP of 0.956; Hara *et al*[29], with 29 SNPs for a breed native to Japan reported a CIP of 2.73 x 10^{-12} and a CEP of 0.96929 to 0.99693. In other related studies, Werner *et al*[24] published a CEP greater than 0.9999 and a CIP of 1 x 10^{-13} with 37 SNPs. Fernández *et al*[17], in Angus with an arrangement of 116 SNPs, reported combined non-exclusion probabilities (CNEP = 1 – CEP) in the range of 2.1 x 10^{-4} to 1.4 x 10^{-9}, as well as CIP of 4.1 x 10^{-15}. Panetto *et al*[16], for the Sindhi breed from Brazil, with 71 SNPs where LAF was higher than 0.35, published CNEP of 1 x 10^{-8}. Zhang *et al*[18], in Simmental cattle with 50 SNPs and LAF greater than 0.40, reported CEP greater than 0.9989; Hu *et al*[19], in crossbred cattle from China, with 50 SNPs where the average LAF value was 0.43, obtained CEP from 0.99797 to 0.999999.

Table 3: Exclusion and identity probability values, obtained with 50 SNPs within the total panel selected by breed

Breed	CEP1	CEP2	CEP3	SNPi	CIP1	CIP2
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

CEP1= combined exclusion probability, with a candidate parent and another known parent. CEP2= combined exclusion probability, given a candidate parent and progeny. CEP3= combined exclusion probability with two candidate parents. CIP1= combined identity probability for two individuals taken at random. CIP2= combined identity probability, for two full siblings taken at random. SNPi= number of SNPs required to obtain a value greater than 0.999999 in the probabilities of exclusion.

For Brangus, Hereford and Limousin cattle, the number of SNPs that make up the panel for paternity tests was greater than 100; selected based on the criteria associated with genetic variability and population structure, with values of exclusion probability greater than 0.999999 and identity probability below 6.6 x 10^{-12}.

Acknowledgements

Thanks to the National Association of Brangus and Limousin Cattle Breeders; as well as the Mexican Hereford Association, for providing the database of the present study. Thanks to the National Council for Science and Technology for the scholarship for postgraduate studies provided.

All authors declare that there is no conflict of interest.

Literature cited:

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 Iris 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. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics 2012;28:2537-2539.