Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Genome-wide estimates of effective population size in the Spanish Holstein population

S. T. Rodríguez-Ramilo[1], J. Fernández[1], M. A. Toro[2], D. Hernández[3] and B. Villanueva[1]
[1]Dpto. Mejora Genética Animal, INIA, [2]Dpto. Producción Animal, ETS Ingenieros Agrónomos,
[3]Dpto. Técnico CONAFE, Madrid, Spain

ABSTRACT: The estimation of inbreeding and coancestry in the Holstein cattle breed has great interest given the rates at which both parameters have increased as a consequence of the intense selection practiced in this breed. Traditionally, inbreeding and coancestry coefficients have been calculated from genealogical records. However, the development of genome-wide single nucleotide polymorphisms has increased the interest of calculating these coefficients from genomic information. In this study, genomic estimates of inbreeding, coancestry and effective population size were compared with genealogical estimates in the Spanish Holstein population. Genomic estimates of inbreeding included those obtained on a marker-by-marker basis and those based on runs of homozygosity. The results indicate that the regression coefficient of genomic inbreeding on genealogical inbreeding was higher than the regression coefficient of genealogical inbreeding on genomic inbreeding. Similar results were also observed for coancestry. In addition, inbreeding estimates based on runs of homozygosity tend to reflect identity by descent status better than estimates obtained on a marker-by-marker basis. Estimates of effective population size obtained from genomic and genealogical information were consistent and ranged from about 66 to 79. These values emphasize the need of controlling the rate of coancestry and inbreeding in Holstein selection programmes.
Keywords: dairy cattle; effective population size; runs of homozygosity

Introduction

Practices of intense selection of sires, artificial insemination, and embryo transfer have impacted heavily on the effective population size, levels of genetic variation and inbreeding of the Holstein cattle breed (Purfield et al. (2012)). Traditionally, inbreeding and coancestry coefficients have been calculated from genealogical records. However, the current availability of thousands of single nucleotide polymorphisms (SNPs) genotypes opens up new opportunities for obtaining more accurate estimates (Bjelland et al. (2013)). In particular, genomic approaches may help to avoid several drawbacks of using genealogical information (Keller et al. (2011)). First, genealogical inbreeding and coancestry describe identity by descent (IBD) status with respect to an arbitrarily defined founder population in which individuals are assumed to be unrelated and non-inbred. Consequently, this approach fails to capture the influence of relatedness among founders. Second, genealogical inbreeding and coancestry are expected proportions of the genome that are IBD and do not take into account the stochastic nature of recombination. Finally, genome-wide data have the advantage of allowing us the determination of homozygosity within specific genomic regions and across chromosomes.

The objective of this study was to obtain genomic estimates (based on SNP data) of inbreeding, coancestry and effective population size in the Spanish Holstein population. Genomic estimates were then compared with those obtained from genealogical information.

Materials and Methods

Genomic and genealogical data. Genomic information from 11,135 individuals belonging to the Spanish Holstein population was analyzed in this study. These animals were genotyped with the Illumina BovineSNP50 BeadChip (versions v1 or v2). Only SNPs common to both chip versions were selected for the analysis (52,340). Unmapped SNPs (523) and those mapped on chromosomes X or Y according to the UMD 3.0 genome assembly were excluded (1,056). In addition, SNPs (14,068) and individuals (566) with more than 5% missing genotypes were discarded. The final dataset included 36,693 SNPs and 10,569 individuals.

The genealogy constructed with all known ancestors of genotyped individuals comprised 35,473 animals. The generation interval inferred from this genealogical data set was 4.2 years.

Estimates of coancestry and inbreeding. Two types of estimates based on SNP data were evaluated: genomic estimates calculated on a marker-by-marker basis and genomic estimates based on runs of homozygosity (ROH). They were compared with estimates obtained from genealogical data.

Estimates obtained on a marker-by-marker basis. Following Malécot (1948), the molecular coancestry, f_{Mij}, between individuals i and j, is the probability that two alleles at a given locus taken at random from each individual are equal (identical by state, IBS). Analogously, F_{Mi} (the molecular inbreeding of individual i) is the probability that the two alleles carried by this individual at a given locus are IBS. The molecular coancestry between individuals i and j was calculated as

$$f_{Mij} = (1/S)\sum_{s=1}^{S}\left[\left(\sum_{k=1}^{2}\sum_{m=1}^{2} I_{sk(i)m(j)}\right)/4\right],$$

where S is the number of SNPs and $I_{sk(i)m(j)}$ is the identity of the k^{th} allele from individual i with the m^{th} allele from individual j at SNP s, that takes a value of 1 if alleles are identical and 0 otherwise. The inbreeding coefficient of individual i was calculated as $F_{Mi} = 2f_{Mii} - 1$.

Estimates based on ROH. Runs of homozygosity are long, uninterrupted stretches of homozygous genotypes that enable reliable estimation of levels of inbreeding (Keller et al. (2011)). The inbreeding estimator, F_{ROH}, refers to the proportion of the genome that is in ROHs. For individual i, F_{ROHi} was calculated as

$$F_{ROHi} = (1/l_g)\sum_{k} length(ROH_k),$$

where ROH_k is the k^{th} ROH in individual i's genome, and l_g is the length of the genome in base pairs. The criteria used for defining a ROH were as follows: (i) a maximum of two missing genotypes and one heterozygous genotype within a particular ROH were permitted; (ii) the minimum density was 1 SNP per 100 Kb; (iii) the maximum distance allowed between two consecutive homozygous SNPs in a run was 1 Mb; and (iv) the minimum size that constituted a ROH was 1 Mb.

Genealogical estimates. The genealogical coancestry (f_{Gij}) and inbreeding (F_{Gi}) coefficients for the genotyped individuals were calculated going back five generations of ancestors. This data set included 31,203 animals. Estimates of both coefficients were obtained using the algorithm of Meuwissen and Luo (1992), which is implemented in the software Pedig (Boichard (2002)).

Correlation and regression between genomic and genealogical coefficients. Pearson´s correlation coefficients between SNP-based and genealogy-based estimates were calculated. In addition, the regression of $Ln(1 - F_G)$ on $Ln(1 - F_M)$ (or $Ln(1 - F_{ROH})$) and that of $Ln(1 - F_M)$ (or $Ln(1 - F_{ROH})$) on $Ln(1 - F_G)$ were performed. Equivalent regressions for coancestry were also carried out.

Rates of change in coancestry and inbreeding, and effective population size. Rates of change in (genomic and genealogical) inbreeding per year ($\Delta F_{M(y)}$, $\Delta F_{ROH(y)}$ and $\Delta F_{G(y)}$) were computed by regressing the natural logarithm of $(1 - F)$ for each individual on year of birth. The slopes of these regressions are approximately equal to $-\Delta F_{M(y)}$, $-\Delta F_{ROH(y)}$ and $-\Delta F_{G(y)}$. Rates of change in inbreeding per generation (ΔF_M, ΔF_{ROH} and ΔF_G) were calculated as $L\Delta F_{M(y)}$, $L\Delta F_{ROH(y)}$ and $L\Delta F_{G(y)}$ where L is the generation interval. Finally, the effective population size was estimated from the rate of change in inbreeding per generation as $N_{eF\_M} = 1/2\Delta F_M$, $N_{eF\_ROH} = 1/2\Delta F_{ROH}$ and $N_{eF\_G} = 1/2\Delta F_G$. Rates of change in coancestry per year ($\Delta f_{M(y)}$ and $\Delta f_{G(y)}$), and per generation (Δf_M and Δf_G) and the effective population size ($N_{ef\_M} = 1/2\Delta f_M$ and $N_{ef\_G} = 1/2\Delta f_G$) were also computed.

Results and Discussion

Genomic coefficients of inbreeding and coancestry were much higher than genealogical coefficients because the later assume no homozygosity in the founder population. Mean values (± standard deviation) for F_M, F_{ROH} and F_G were 0.65 (± 0.01), 0.12 (± 0.03) and 0.04 (± 0.03), respectively. Mean values (± standard deviation) for f_M and f_G were 0.64 (± 0.01) and 0.04 (± 0.02), respectively. Many estimators have been proposed to correct molecular-based coefficients for the homozygosity existing in the base population (Toro et al. (2011)). However, these methods are not accurate when, as usual, the allele frequencies in the base population are unknown (Toro et al. (2002)).

Table 1 shows the regression and correlation coefficients between different coancestry and inbreeding estimates. Pearson's correlation coefficient between f_M and f_G (0.73) was substantially higher than that between F_M and F_G (0.51). Also, the regression coefficient of f_M on f_G and that of f_G on f_M were higher than the corresponding coefficients involving inbreeding coefficients. Table 1 also shows that the regression coefficient of f_M on f_G was higher (b = 0.83) than that of f_G on f_M (b = 0.66). Similar results were also observed for F_M (or F_{ROH}) and F_G.

Table 1. Regression coefficients (*b*) and correlations (*r*) between different coancestry (*f*) and inbreeding (*F*) estimates.

Regression of			*b*	*r*
$Ln(1-f_G)$	on	$Ln(1-f_M)$	0.66	0.73
$Ln(1-f_M)$	on	$Ln(1-f_G)$	0.83	
$Ln(1-F_G)$	on	$Ln(1-F_M)$	0.37	0.51
$Ln(1-F_M)$	on	$Ln(1-F_G)$	0.70	
$Ln(1-F_G)$	on	$Ln(1-F_{ROH})$	0.44	0.59
$Ln(1-F_{ROH})$	on	$Ln(1-F_G)$	0.79	
$Ln(1-F_{ROH})$	on	$Ln(1-F_M)$	0.88	0.91
$Ln(1-F_M)$	on	$Ln(1-F_{ROH})$	0.93	

f_G: genealogical coancestry; f_M: SNP-by-SNP-based coancestry; F_G: genealogical inbreeding; F_M: SNP-by-SNP-based inbreeding; F_{ROH}: ROH-based inbreeding

It must be highlighted that the correlation and regression coefficients involving F_{ROH} and F_G were higher than those involving F_M and F_G (Table 1). This indicates that F_{ROH} tends to reflect IBD better than F_M as shown by Keller et al. (2011) given that long homozygous segments are only expected if both copies have been inherited from a common ancestor.

The highest correlation and regression coefficients were those between F_M and F_{ROH}. The moderate correlation between F_{ROH} and F_G showed in Table 1 (0.59) is in agreement with the results of Ferenčaković et al. (2012). These authors reported correlations between F_{ROH} and F_G ranging from 0.50 to 0.72 when analyzing 1,421 bulls from four different cattle breeds. In a posterior study, Ferenčaković et al. (2013) indicated that correlations between F_{ROH} and F_G are similar when using the 50K chip (estimates from 0.62 to 0.77) than when using the High Density cattle panel (estimates from 0.61 to 0.75).

Table 2 shows the rates of change in (genomic and genealogical) coancestry and inbreeding per year and per generation, and the effective population size estimated from both rates. The rates of change in coancestry were quite similar when calculated using genomic or genealogical data. Consequently, the effective population sizes estimated from Δf_M and Δf_G were close (70 and 66, respectively).

Table 2. Rates of change in coancestry ($\Delta f_{(y)}$) and inbreeding ($\Delta F_{(y)}$) per year and per generation (Δf and ΔF) using different sources of information. Effective population sizes obtained from Δf (N_{ef}) and from ΔF (N_{eF}) are also presented.

	Information used		
		Genomic	
	Genealogical	SNP-by-SNP	ROH
$\Delta f_{(y)}$	0.0018	0.0017	-
Δf	0.0076	0.0071	-
N_{ef}	66.1	70.0	-
$\Delta F_{(y)}$	0.0015	0.0016	0.0017
ΔF	0.0063	0.0067	0.0071
N_{eF}	79.4	74.4	70.0

Rates of change in genomic and genealogical inbreeding were also similar and they were both slightly lower than the corresponding rates of coancestry. Consequently, estimates of effective population size obtained from rates of inbreeding (74 and 79 when using ΔF_M and ΔF_G, respectively) were slightly higher than estimates obtained from rates of coancestry. This could be a reflection of the avoidance of matings between relatives within this population.

Estimates of effective population size obtained here are close to those previously published for other Holstein populations using genealogical data. Parland et al. (2007) determined the effective population size for the Irish Holstein population to be 75. Koenig and Simianer (2006) estimated a lower effective population size for the German Holstein population (52). These values reinforce the need of controlling the rate at which coancestry and inbreeding increase in Holstein selection programmes.

Conclusion

Our results indicate that the regression coefficient of genomic inbreeding (coancestry) on genealogical inbreeding (coancestry) is higher than the regression coefficient of genealogical inbreeding (coancestry) on genomic inbreeding (coancestry). In addition, inbreeding ROH-based estimates tend to reflect IBD status better than inbreeding estimates obtained on a marker-by-marker basis. Estimates of effective population size, obtained from genomic and genealogical information, ranged from about 66 to 79. These estimates are of the same order of magnitude to those estimated in other Holstein populations and strengthen the need of controlling the rate of coancestry and inbreeding in current Holstein selection programmes.

Acknowledgements

We are very grateful to the EuroGenomics Consortium and Conafe for providing genotypes and genealogical data used in this study. This study was supported by the Ministerio de Economía y Competitividad, Spain (grant CGL2012-39861-C02-02).

Literature Cited

Bjelland, D. W., Weigel, K. A., Vukasinovic, N. et al. (2013). J. Dairy Sci. 96: 4697-4706

Boichard, D. (2002). Proceedings of the 7th WCGALP, Montpellier, FR, paper 28-13

Ferenčaković, M., Hamzić, E., Gredler, B. et al. (2012). J. Anim. Breed. Genet. 130: 286-293

Ferenčaković, M., Sölkner, J., and Curik, I. (2013). Gen. Sel. Evol. 45: 42

Keller, M. C., Visscher, P. M., and Goddard, M. E. (2011). Genetics, 189: 237-249

Koenig, S. and Simianer, S. (2006). Lives. Sci. 103: 40-53

Malécot, G. (1948). Les mathématiques de l'hérédité. Paris: Masson & Cie

Meuwissen, T. H. and Luo, Z. (1992). Gen. Sel. Evol. 24: 305-313

Parland, S. Mc., Kearney, J. F., Rath, M., et al. (2007). J. Anim. Sci. 85: 322-367

Purfield, D. C., Berry, D. P., McParland, S., et al. (2012). BMC Genet. 13: 70

Toro, M. A., Barragán, C., Óvilo, C., et al. (2002). Conserv. Genet. 3: 309-320

Toro, M. A., García-Cortés, L. A., and Legarra, A. (2011). Gen. Sel. Evol. 43: 27