Measuring inbreeding in a closed pig strain from high-density SNPs genotypes

L. Silió[1], A. Fernández[1], A. Mercadé[2], P. Martin-Palomino[3], M.A. López[1], J. Rodrigáñez[1] and C. Ovilo[1]

Introduction

Inbreeding plays a key role in genetics applied to livestock production, and it is usually measured from the pedigree records of a population using the inbreeding coefficient *F*, defined as the probability that the two homologous alleles in an individual are identical by descent. Instead of calculating genealogical inbreeding coefficients, several studies in farm animals have reported the use of microsatellite-derived metrics to infer relative inbreeding from the homozygosity of the markers analyzed (Slate *et al*., 2004; Alves *et al.*, 2008). Their main goal was to determine if ranking individuals for molecular homozygosity is equivalent to ranking them for the genealogical inbreeding coefficient *F*, although their results were not totally satisfactory due to the low variation of *F* values. The relationship between homozygosity and *F* depends on the number and heterozygosity of markers but it mostly depends on the variance of *F* (Pemberton, 2004).

Single nucleotide polymorphisms (SNPs) have become a useful type of marker for diverse purposes as automated genotyping systems have been developed that yield accurate genotypes (Rohrer *et al.*, 2007). Recently, a high-density porcine SNP genomic assay is commercially available (Illumina Porcine SNP60 BeadChip). This assay is an excellent tool for studies of pig genetic variability because it contains probes to genotype thousands of biallelic SNPs covering the whole pig genome (Ramos *et al*. 2009). In the present study, 65 individuals from a closed strain of Iberian pigs, with 26 overlapped generations of pedigree completely known, have been genotyped for 62,163 SNPs included in the Beadchip. Their molecular inbreeding values (F_{SNP}) calculated from these genotypes have been compared with the inbreeding coefficients calculated from either the complete pedigree (F_{26G}) or the pedigree from the last five generations (F_{5G}). In order to check cheaper alternatives and the consistency of results for the different autosomes, the relationships between the correlation $r(F_{26G}, F_{SNP})$ and the number and genome distribution of SNPs typed have also been examined.

Material and methods

Animals. The Torbiscal strain was founded in 1963 from the mixture of four ancient Iberian pig populations kept genetically isolated since 1945. The complete genealogy of all the animals is available, with 1,695 entries animal-sire-dam and 25.8 generations from the founders to the breeding animals genotyped in the present study. These animals include seven boars and six sows. Moreover, 52 progenies born in six litters were also genotyped: three of them (27 piglets) proceed from full-sibs matings and the other three (25 piglets) from minimum coancestry matings.

[1]Departamento de Mejora Genética Animal, INIA, Madrid, Spain
[2]Department de Ciencia Animal i del Aliments, UAB, Barcelona, Spain
[3]CIA 'Dehesón del Encinar', Toledo, Spain

Genotyping. The 65 animals were genotyped with the Porcine SNP60 BeadChip (Illumina) using the manufacturer recommendations and visualized and analyzed with the GenomeStudio software (Illumina). The genotyping of 2,268 out of the 62,163 SNPs was not feasible for technical reasons.

Statistical analysis. Genealogical inbreeding coefficients of the genotyped pigs were calculated tracing the pedigree back 26 generations to the founder animals (F_{26G}), or alternatively tracing the pedigree back only 5 generations to common ancestors (F_{5G}). The molecular inbreeding or homozygosity was calculated as the proportion of homozygous SNP (identical by state), although only 40,200 out of the 59,895 available SNPs are useful for these calculations (F_{60KSNP}), because of 18,386 are not polymorphic in the Iberian pig breed and other 1,309 SNPs map on chromosome X. Other molecular homozygosity values were calculated from diverse panels of SNPs (from n = 24 to 384) selected by their intermediate frequencies or for their location on each one of the 18 autosomes.

Results and discussion

The main statistics of both genealogical and molecular inbreeding are given in Table 1. F values calculated from the last five generations (F_{5G}) are remarkably well correlated with those calculated from the complete available pedigree (F_{26G}), because of recent inbreeding events have a larger influence on an individual's inbreeding coefficient relative to events deeper in the pedigree (Balloux *et al*., 2004). Moreover, both genealogical coefficients (F_{26G} and F_{5G}) present a great dispersion due to the population mating system combining full-sibs and minimum coancestry matings. This family structure increases the variance of F genealogical estimates and therefore their correlation with other measures of inbreeding (Pemberton, 2004).

Table 1. Statistics of coefficients of genealogical inbreeding of 65 analyzed pigs calculated from the complete pedigree (F_{26G}) or from the last five generations (F_{5G}), and of molecular inbreeding calculated from 62,163 SNPs (F_{60KSNP}) or from a selected set of 192 SNPs (F_{192SNP}), and their respective correlation coefficients (r). All the r values are significant at $P < 0.0001$

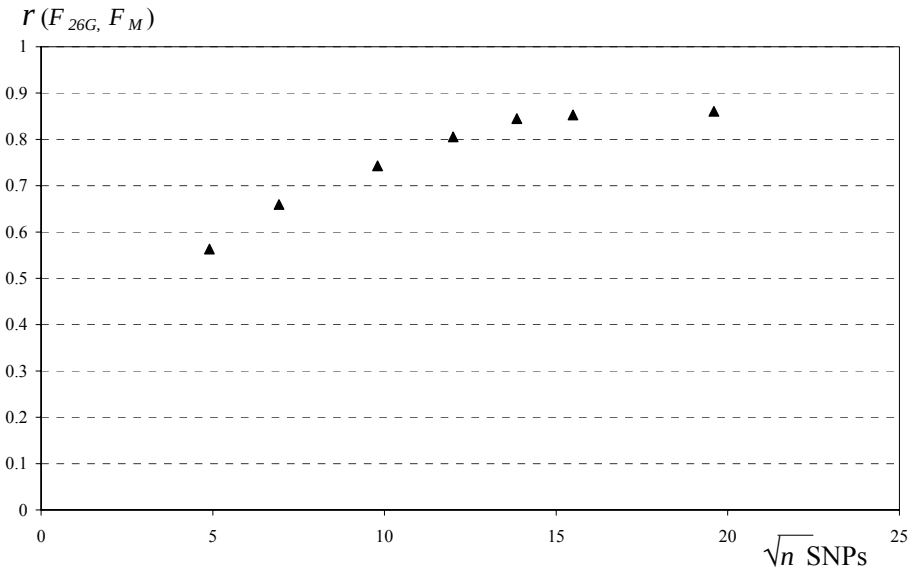
	Mean	Standard Deviation	Minimum/ Maximum	Correlation Coefficient (r)		
				F_{5G}	F_{60KSNP}	F_{192SNP}
F_{26G}	0.258	0.101	0.166/0.382	1.000	0.917	0.845
F_{5G}	0.111	0.121	0.000/0.260		0.917	0.846
F_{60KSNP}	0.769	0.033	0.721/0.835			0.890
F_{192SNP}	0.530	0.085	0.377/0.721			

The values for the molecular inbreeding (F_{60KSNP} and F_{192SNP}) are higher than for the genealogical ones (F_{26G} and F_{5G}). These biased values may be explained by the relationship between genealogical (F) and molecular inbreeding (F_{SNP}): $1 - F_{SNP} = (1 - \Sigma\, p_i^2)\,(1 - F)$, where $(1 - F)$ and $(1 - F_{SNP})$ are the heterozygosities by descent and by state in the current

population, and $(1 - \Sigma\, p_i^2)$ is the expected heterozygosity by state in the founder population (Alves *et al*., 2008). This measure of the variability of the genotyped SNPs in the founder population may be estimated applying the quoted equation to each genotyped individual. Two similar values of $(1 - \Sigma\, p_i^2)$ were obtained by average (0.312) or by regression (0.300).

The mean values of molecular inbreeding coefficients ($F_{n\text{SNP}}$) calculated from different sets of selected SNPs ranged from $F_{48\text{SNP}} = 0.510$ to $F_{384\text{SNP}} = 0.538$. All these values were lower than the obtained from the complete 60K Beadchip: $F_{60\text{KSNP}} = 0.769$ (Table 1). This bias may be explained because of the SNPs included in these cheaper chips were selected among those of maximum mean heterozygosity (≈ 0.50) and spaced along the different autosomes. Moreover, a moderate number of these SNPs ($n = 192$) may provide –for this particular coancestry structure – inbreeding coefficients with a strong correlation with those based on the complete or partial pedigree. According to Chakraborty (1981) the expected correlation between the heterozygosity of the genome of N unlinked loci and of a sample of n loci is $(n/N)^{-1/2}$. The proportionality to $n^{-1/2}$ of the correlation between genealogical and molecular inbreeding has been examined in our data for these sets of SNPs. The results confirm this proportionality from 24 to 192 selected SNPs, but the use of larger sets of SNPs ($n = 240$ or 384) only slightly improve the correlation because the mandatory addition of markers with heterozygosity lower than 0.50 (Figure 1).

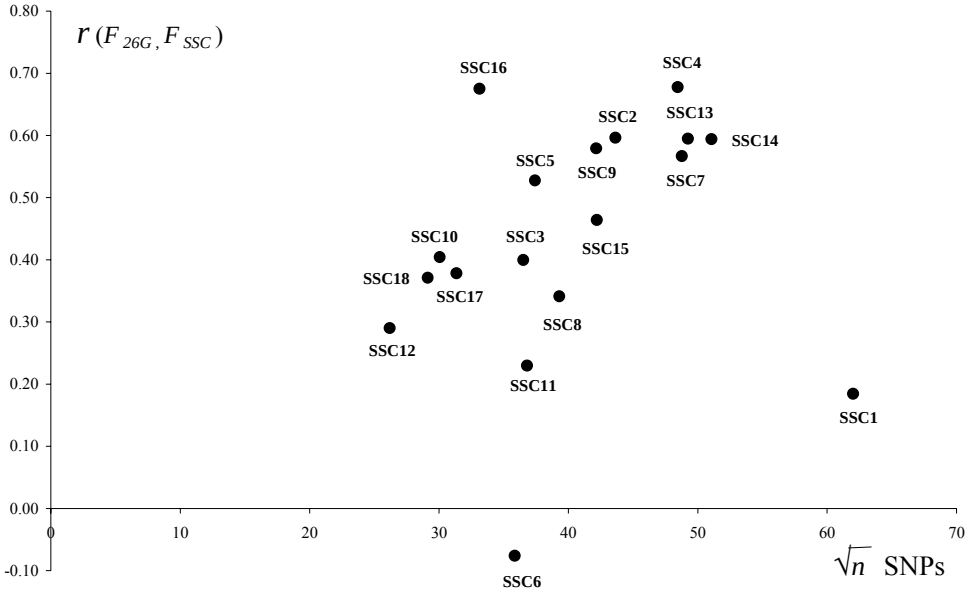
Figure 1. Estimated correlation coefficients between genealogical inbreeding based on the complete pedigree and molecular inbreeding r ($F_{26\text{G}}$, $F_{n\text{SNP}}$) for diverse sets of selected SNPs ($n = 24$, 48, 96, 144, 192, 240 and 384) as a function of the square root of the number of SNP genotyped (n).



The known chromosomal location of an important number of SNPs allows calculate the coefficient of molecular inbreeding (F_{SSC}) separately for each one of the autosomes. The

chromosomal F_{SSC} values present a range of variation from 0.722 (SSC1) to 0.824 (SSC2). The correlation values between chromosomal and genealogical inbreeding are reported in Figure 2. For 14 out of the 18 autosomes, the respective correlation values r (F_{26G}, F_{SSC}) clearly depend on the square root of the number of SNP, but the other four autosomes present outlier r values. The low correlation values of three of them (SSC1, SSC6, and SSC11) may be attributed to the low standard deviation among individuals of the molecular inbreeding for these chromosomes. The research on genetic causes of this lower variation exceeds the goals of the present study, and it should be carried out soon later.

Figure 2. Estimated correlation coefficients between genealogical and molecular chromosomal inbreeding r (F_{26G}, F_{SSC}) as a function of the square root of the number of SNP genotyped in each chromosome (n).



References

Alves, E., Barragán, C. and Toro, M.A. (2008). *Spanish J. Agr. Research,* 6:248–251.

Balloux, F., Amos, W. and Coulson, T. (2004). *Molec. Ecol*. 13:3021–3031.

Chakraborty, R. (1981). *Genetics,* 98:461–466.

Pemberton, J. (2004). *Trends Ecol. Evol.,* 19:613–615.

Ramos, A.M., Crooijmans R.P.M.A., Affara N.A. *et al.* (2009). *PLoS One*, 4:e6524.

Rohrer, G.A., Freking, B.A. and Nonneman, D. (2007). *Anim. Genet.,* 38:253–258.

Slate, J., David, P., Dodds, K.G. *et al.* (2004). *Heredity,* 93:255–265.