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

Inbreeding by Pedigree and Genomic Markers in Selection Lines of Pigs

Y. Zhang[1,2], J. M. Young[1], C. Wang[1], X. Sun[1], A. Wolc[1] and J. C. M. Dekkers[1].
[1]Iowa State University, Ames, USA, [2]China Agricultural University, Beijing, China.

ABSTRACT: Dense SNP data of 2,358 pigs were analyzed to quantify inbreeding across nine generations in two Yorkshire selection lines that were divergently selected for residual feed intake. Alternative genomic inbreeding estimates were compared with pedigree-based inbreeding. In general, genomic markers traced the same trends in inbreeding as pedigree, although the observed rate of inbreeding was lower when based on genetic markers. Higher within-generation variation was observed for the marker-based estimates of inbreeding, implicating that the markers can capture Mendelian sampling variance and reveal the 'realized' homozygosity in the genome. The marker-based estimates of inbreeding were highly or moderately correlated with each other (0.74-0.98), but less correlated with pedigree-based estimates of inbreeding (0.57-0.69).
Keywords: genomic inbreeding; residual feed intake; runs of homozygosity

Introduction

Inbreeding results in an increase in homozygosity, which leads to three negative effects: increased prevalence of deleterious recessive alleles, a reduction in phenotypic values for some traits, i.e. inbreeding depression, and a reduction in genetic variance. In farm animal populations, inbreeding can accumulate rapidly due to intense selection and small population size.

The inbreeding coefficient (F), defined as the probability that the two alleles at a random locus in an individual are identical by descent, has been a widely used measure of inbreeding. Inbreeding has traditionally been computed based on pedigree, although its reliability is often affected by problems such as incomplete or incorrect pedigree records. In recent years, the availability of dense SNPs has facilitated the quantification of inbreeding by genomic markers in farm animals (e.g. VanRaden et al. (2011)).

Genomic inbreeding can be described as individual genome-wide homozygosity-by-descent. There are two categories of genomic inbreeding measures based on genome-wide SNPs. One is based on marker-by-marker estimates, such as the diagonal elements of the genomic relationship matrix (GRM) (VanRaden (2008)), the canonical estimate based on excess SNP homozygosity in PLINK (Purcell et al. (2007)), the low sampling error estimate proposed by Yang et al. (2010)), and molecular coancestry estimates (Toro et al. (2011)). The other category is to detect runs of homozygosity (ROH), which are defined as stretches of continuously homozygous SNPs that span a certain minimum length (e.g. 0.5 Mb, 1.5 Mb, 5 Mb).

The aim of this work was to quantify inbreeding in two Yorkshire lines that were selected for high and low residual feed intake (RFI), to monitor inbreeding trends across generations in these lines, and to compare pedigree inbreeding with alternative genomic inbreeding estimates.

Materials and Methods

Population. The pedigree and genotype data were from two lines of Yorkshire pigs which underwent bi-directional, single-trait selection for RFI at Iowa State University (Cai et al. (2008)). The two lines were started from split litters in generation 0. The low RFI line was selected for reduced RFI (increased efficiency) for 8 generations. The high RFI line was randomly mated until generation 4 and then selected for increased RFI (reduced efficiency) since generation 5.

Genotypes, quality control and imputation. A total of 2,380 pigs were genotyped on the Illumina Porcine SNP60 BeadChip. Data quality control was conducted using PLINK (Purcell et al. (2007)), with the criteria for exclusion including >10% missing genotypes by individual and by SNP, SNP minor allele frequency (MAF) <1%, and Hardy-Weinberg equilibrium P value $<10^{-19}$. In addition, one outlier individual with a high proportion of heterozygous sites and 17 samples that failed parentage testing were removed. Finally, 2,358 animals and 48,156 SNPs were used for further analysis (Table 1). Missing genotypes were imputed by AlphaImpute which uses segregation analysis and haplotype library imputation (SAHLI) to impute alleles and genotypes in a pedigreed population (Hickey et al. (2012)).

Table 1. The number of genotyped animals in low (L) and high (H) RFI lines across 9 generations.

Generation	L	H	Total
0	118	19	137
1	67	0	67
2	167	23	190
3	66	20	86
4	185	101	286
5	191	227	418
6	136	115	251

7	185	170	355
8	315	253	568
Total	1,430	928	2,358

Estimates of F. Six alternative measures of inbreeding were calculated for each individual.

1. F_{PED} is the pedigree-based F calculated in CFC (Sargolzaei et al. (2006)).

2. F_{GRM} is the diagonal elements of the genomic relationship matrix proposed by VanRaden (2008):

$$F_{GRM} = \frac{\sum_{i=1}^{N}(x_i - 2p_i)^2}{\sum_{i=1}^{N}(2p_iq_i)} - 1$$

where x_i is the genotype for the i^{th} SNP in all individuals coded as 0, 1, or 2, p_i is the allele frequency for the i^{th} SNP in the base population (generation 0 in this study), q_i=1-p_i, N is the number of SNPs. Allele frequencies in the base population were estimated as in VanRaden (2008), using the algorithm of Gengler et al. (2007).

3. $F_{GRM0.5}$ is the same as F_{GRM}, except setting all allele frequencies to 0.5. VanRaden et al. (2011) demonstrated that using frequencies of 0.5 is advantageous because of the challenges associated with estimation of allele frequencies in the base population.

4. F_{Yang} is an alternative genomic inbreeding estimate with low sampling error proposed by Yang et al. (2010):

$$F_{Yang} = \sum_{i=1}^{N} \frac{x_i^2 - (1+2p_i)x_i + 2p_i^2}{2p_iq_i}$$

5. F_{HOM} is a simple genomic inbreeding measure based on the percentage of homozygous genotypes across all SNPs:

$$F_{HOM} = N_{\text{hom}} / N$$

6. F_{ROH} is the proportion of the genome that is in runs of homozygosity (RHOs).

$$F_{ROH} = \frac{\sum_k length(ROH_k)}{L}$$

where ROH_k is the k^{th} ROH discovered in an animal and L is the length of the genome. Here we used the total length of 18 autosomes (swine genome build 10.2). ROHs were detected using PLINK. Since high linkage disequilibrium (LD) within dense SNP regions can lead to chance (non-IBD) ROH segments (Howrigan et al. (2011)), LD-pruning was performed before ROH discovery. A 50-SNPs window was slid along the chromosome with step size 5 and SNPs with r^2 >0.5 with all other SNPs in the window were removed, which resulted in 14,366 SNPs remaining. ROHs were detected with the following criteria: a minimum ROH length of 5,000 kb; at least 10 SNPs in the ROH; at most 500 kb/SNP in the ROH; maximum gap of 1,000 kb between two consecutive SNPs within the ROH; at most 1 missing SNP allowed in the ROH.

Results and Discussion

Selection lines offer an opportunity to assess marker-based estimates as alternative indicators of inbreeding. In the current study, all estimates of inbreeding, except F_{GRM} and F_{Yang}, ranged from 0 to 1. In early generations (from 0 to 3), negative values were observed for F_{GRM} and F_{Yang}. Scales differed between alternate estimates of inbreeding. As shown in Table 2, the pedigree-based estimate of inbreeding (F_{PED}) showed much lower variance, which can be explained by it being an expectation of genome-wide IBD based on pedigree relationships, while marker-based estimates of inbreeding potentially reveal the individual 'realized' homozygosity.

Table 2. Average within-generation standard deviation (w/n-Gen. SD), regression of inbreeding coefficients (intercept and b) on generation for the low (L) and high (H) RFI lines.

	w/n-Gen. SD		Rate of Inbreeding			
	L	H	b_L	$Intercept_L$	b_H	$intercept_H$
F_{PED}	0.010	0.018	0.017	-0.004	0.013	-0.001
F_{GRM}	0.045	0.054	0.014	0.020	0.009	0.047
$F_{GRM0.5}$	0.031	0.035	0.010	0.266	0.008	0.264
F_{Yang}	0.041	0.050	0.015	-0.014	0.012	-0.005
F_{HOM}	0.016	0.019	0.005	0.640	0.003	0.642
F_{ROH5M}	0.023	0.027	0.010	0.058	0.006	0.058

As shown in Table 3, correlations between pedigree-based and marker-based estimates of inbreeding were relatively low (0.57-0.69). Correlations among F_{HOM}, F_{Yang} and $F_{GRM0.5}$ were high (0.90-0.98), while F_{GRM} and F_{ROH} were only moderately correlated with the other estimates. It is worth noting that F_{HOM}, F_{ROH}, and $F_{GRM0.5}$ can be directly calculated without estimates of allele frequencies, which are required for F_{GRM} and F_{Yang}.

Table 3. Pearson correlations among alternative estimators of individual inbreeding coefficients.

	F_{PED}	F_{GRM}	$F_{GRM0.5}$	F_{Yang}	F_{HOM}
F_{GRM}	0.57				
$F_{GRM0.5}$	0.63	0.79			
F_{Yang}	0.64	0.93	0.92		
F_{HOM}	0.58	0.78	0.98	0.90	
F_{ROH}	0.69	0.74	0.87	0.84	0.84

F_{HOM} and F_{ROH} are sensitive to the user defined parameters in the analysis. E.g., for F_{HOM}, increasing the minimum MAF from 0.01 to 0.05 led to a drop in the

average inbreeding for generation 0 in the low RFI line from 0.63 to 0.61 (data not shown).

Although a recent study (Keller et al. (2011)) has shown that F_{ROH} is preferable to F_{PED} and to marker-by-marker estimates of inbreeding for the detection of both recent and historic inbreeding, a limitation to its application is that it is affected by various tuning parameters. In this study, we detected homozygous segments as long as 5,000 bp as ROHs after LD-pruning. Long ROHs better reflect recent inbreeding (Keller et al. (2011)), which may explain the relatively high correlation of F_{ROH} with F_{PED}.

As shown in Figure 1, the average inbreeding level increased by generation. The rate of inbreeding per generation based on pedigree was 0.017 and 0.013 in the low and high RFI lines, respectively. The lower inbreeding rate for the high RFI line likely is due to the fact that selection was at random during generations 0 to 3, although the number of males used for breeding was lower in the high than the low RFI line during those generations. All marker-based estimates of inbreeding had the same trend as F_{PED} over generations (Figure 1) but showed a lower inbreeding rate than F_{PED} (Table 2).

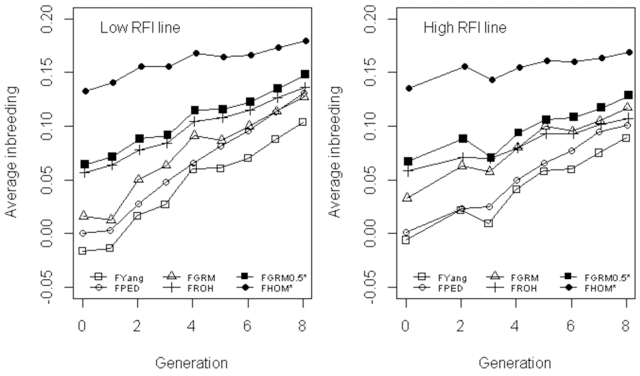

Figure 1. Inbreeding trends in selection lines of pigs measured by pedigree and dense markers[§].

[§] Average estimates of inbreeding of genotyped animals plotted against generations. The original values of $F_{GRM0.5}$ and F_{HOM} were reduced by 0.2 and 0.5 to better fit the figures.

Selection for RFI has been successful for these pig populations (Cai et al. (2010)). However, the small population sizes and intense selection has led to an accumulation of inbreeding, as revealed by both pedigree and genomic markers. Inbreeding control is necessary to limit the possible impact of deleterious alleles, inbreeding depression, and loss of variance. In addition to the traditional strategies such as avoidance of matings between closely related individuals and the introduction of outside boars, using genomic relationships may help to better control genome-based inbreeding in the future.

Conclusions

Genomic markers traced the same trends of inbreeding as pedigree, although the rate of inbreeding was lower for the marker-based estimates. Higher within-generation variation was observed for the marker-based estimates of inbreeding, implicating that the markers can capture the Mendelian sampling variance and reveal the 'realized' homozygosity in the genome. The marker-based estimates of inbreeding were highly or moderately correlated with each other, but less correlated with the pedigree-based estimate of inbreeding. F_{HOM} and $F_{GRM0.5}$ are the most straightforward measures of genomic inbreeding, while estimation of F_{GRM}, F_{Yang} and F_{ROH} are relatively difficult because they require estimates of allele frequencies in the base population or a number of user-defined parameters.

Acknowledgements

Support for development of the ISU RFI lines and genotyping from the National Pork Board, the Iowa Pork Producers Association, Zoetis, the USA Swine Genome Coordinator, the Iowa Experiment Station and Hatch Act and State of Iowa Funds (Project No. 3600), and the Agriculture and Food Research Initiative Competitive Grant no. 2011-68004-30336 from the USDA National Institute of Food and Agriculture.

Literature Cited

Cai, W., Casey, D. S., and Dekkers, J. C. M. (2008). J. Anim. Sci., 86(2), 287-298.

Clark, S. A., Kinghorn, B. P., Hickey, J. M., et al. (2013). Genet. Sel. Evol., 45, 44.

Gengler, N., Mayeres, P., and Szydlowski, M. (2007). Animal, 1: 21-28.

Hickey, J. M., Kinghorn, B. P., Tier, B., et al. (2012). Genet. Sel. Evol., 44: 9.

Howrigan, D. P., Simonson, M. A., and Keller, M. C. (2011). BMC Genomics, 12, 460.

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

Purcell, S., Neale, B., Todd-Brown, K., et al. (2007). Am. J. Hum. Gen., 81, 559-575.

Sargolzaei, M., Iwaisaki, H., and Colleau, J.J. (2006). Proc 8th WCGALP, volumn 27, 28.

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

VanRaden, P. M. (2008). J. Dairy Sci., 91, 4414-4423.

VanRaden, P. M., Olson, K. M., Wiggans, G. R., et al. (2011). J. Dairy Sci., 94, 5673-5682.