Proc. Assoc. Advmt. Anim. Breed. Genet. 22:249-252

DEVIATIONS AROUND KINSHIP EXPECTATIONS AT VARIOUS SNP MARKER DENSITIES IN A POPULATION OF BROILER CHICKEN

A. Reverter[1], Y. Li[1], A. George[2], J. Henshall[3], R. Sapp[3], R. Okimoto[3], R. Hawken[3] and S. A. Lehnert[1]

[1]CSIRO Agriculture & Food, Queensland Bioscience Precinct, Queensland 4067, Australia.
[2]CSIRO Data61, Ecoscience Precinct, Queensland 4102, Australia.
[3]Cobb-Vantress Inc., Siloam Springs, Arkansas 72761-1030, USA.

SUMMARY

We compare four low-density SNP panels containing 151, 400, 1,000 (1K) and 3,000 (3K) SNP selected from a higher density chip of 50K SNP in their ability to correctly infer 7 kinship relationships (from self-self to grand-mother – grand-offspring) in 4,217 commercial broiler chicken. Self–self relationships estimated from the diagonal elements of the genomic relationship matrix (GRM) were symmetric and centred at 1.0 regardless of the panel used. However, genomic relationships for other relationships were centred slightly left to the expected value indicating possible genotype or pedigree errors. Relationships estimated using either the 1K or the 3K SNP panels were almost undistinguishable from those estimated using the whole 50K chip. However, the two lowest density panels produced relationships with long-tailed distributions. We conclude that a SNP panel of 1K SNP is a cost-effective tool to estimate relationships among individuals.

INTRODUCTION

The ability to correctly infer relationships among individuals underpins the utility of SNP genotype data. This ability is of particular relevance in the development of low-density panels for the implementation of cost-effective genomic strategies. Judge *et al*. (2016a) have recently explored the optimal use of low-density SNP panels for breed assignment in Angus and Hereford cattle. The authors conclude that at least 300 to 400 SNP are needed to accurately predict breed proportions. Similarly, working with various cattle and sheep populations Strucken *et al*. (2016), concluded that at least 700 SNP are needed to fully exclude false positives in parentage assignments. Other authors have evaluated the use of low-density panels for imputation to higher density in cattle (Ogawa *et al*. 2016), sheep (Ventura *et al.* 2016), pig (Badke *et al.* 2014), and chicken (Wang *et al*. 2013).

Here we present four low-density SNP panels containing 151, 400, 1,000 and 3,000 SNP and compare them with the higher density chip of 50K SNP based on their ability to estimate relationships in a population of 4,217 commercial broiler chicken from 22 overlapping generations.

METHODS

Animal resources and relationships considered. We used a total of 4,217 broiler chicken (3,139 females and 1,078 males) from 22 overlapping generations of a commercial line of Cobb-Vantress Inc. The birds were selected from a larger population to ensure parents and grandparents contained within the sample had genotypes for ~50,000 (50K) SNP from the high-density Avian chip from Illumina Inc.

In total, there were 795 dams with genotypes, 117 sires with genotypes and 133 grand-dams with genotypes. With these, seven types of animal to animal relationships were explored including (1) self – self (N = 4,217); (2) full-sibs (N = 29,599 pair combinations); (3) Father – offspring (N = 2,915 pairs); (4) Mother – offspring (N = 2,708 pairs); (5) Paternal half-sibs (N = 186,716 pairs); (6) Maternal half-sibs (N = 1,560 pairs); and (7) Grand-mother – grand-offspring (N = 5,327).

Using the Method 1 of VanRaden (2008) we built the genomic relationship matrix (GRM) across the 4,217 birds using the 50K chip, as well as with the four low-density SNP panels described next.

Low-density panels. For the formation of the low-density SNP panels, we developed a 6-component fitness function to be optimised that included (1) minor allele frequency (MAF); (2) equidistance to ensure uniform genome coverage; (3) distance to known gene; (4) significance of the association to feed-related phenotypes; (5) pleiotropy test statistics; and (6) connectivity in a co-association network.

We used simulated annealing for the optimisation process. Simulated annealing (Kirkpatrick *et al*. 1983) is a heuristic search algorithm for global optimization, using iterative random movements to approximate optimum solution and has gained popularity in the context of livestock genetics and genomics including studies with cattle (Schierenbeck *et al*. 2011) and poultry (Chapuis *et al*. 2016).

Initially, three SNP densities were considered: 400 SNP, 1,000 (1K) SNPs and 3,000 (3K) SNPs. Importantly, these panels were nested such that the 400 SNP in the small panel were contained in the 1K SNP of the medium panel, and these were themselves contained in the larger 3K panel.

In addition, a smaller panel of only 151 SNP was developed. This panel was made of SNP (1) in the coding region of genes reported to be of relevance in the feed efficiency literature; (2) significant (P< 0.01) in the GWAS for at least one of seven feed-related phenotypes previously undertaken; and (3) Included in the 3K SNP panel.

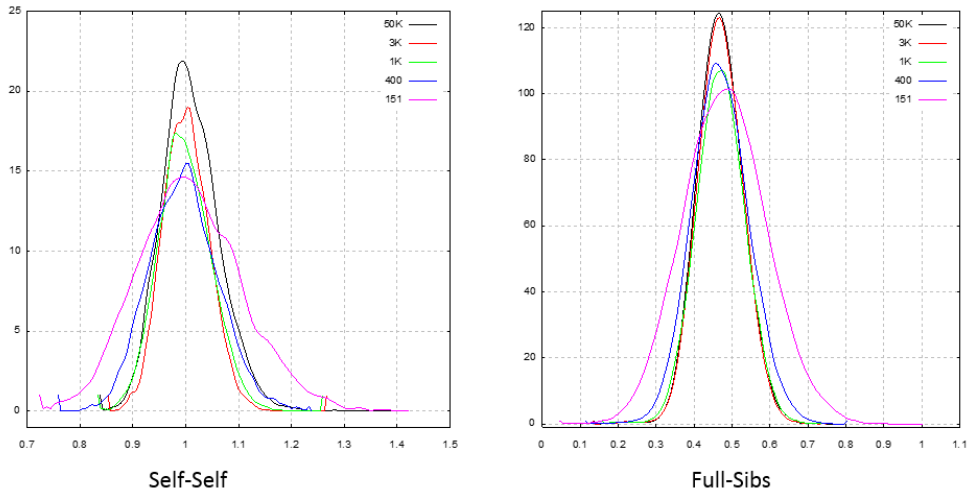

Figure 1. Distribution of genomic relationships for self-self (diagonals elements of the genomic relationship matrix) and full-sibs estimated based on SNP panels of various densities.

RESULTS

Table 1 shows the summary statistics (including mean, standard deviation, minimum and maximum) for genomic relationships estimated using either the high-density 50K SNP chip or the four low-density SNP panels considered is this study and for the seven types of pedigree-based kinships available in our dataset of 4,217 broiler chicken.

Self-self relationships based on the diagonal elements of the GRM were all centred at the expected value of 1. However, the spread was much higher for the panels with only 400 or 151 SNP. Indeed, across all types of relationships considered, the very low density panels of 400 and 151 SNP yielded estimated relationship with higher variation compared to the panels of higher density.

This deviation from expectation is made apparent in Figure 1 for the case of self-self and full-sib relationship in the five SNP panels.

Table 1. Summary statistics for genomic relationships estimated using the high-density 50K SNP chip and four low-density SNP panels for seven types of pedigree-based kinships

Kinship[A]	Panel	No Pairs	Mean	SD	Min.	Max.
SS	50K	4,217	1.009	0.055	0.837	1.404
	3K	4,217	1.001	0.044	0.853	1.267
	1K	4,217	0.999	0.050	0.834	1.259
	400	4,217	0.997	0.066	0.759	1.238
	151	4,217	1.008	0.099	0.723	1.423
FS	50K	29,599	0.469	0.061	0.134	0.769
	3K	29,599	0.470	0.060	0.133	0.749
	1K	29,599	0.470	0.064	0.158	0.739
	400	29,599	0.470	0.075	0.114	0.804
	151	29,599	0.481	0.110	0.045	1.003
FO	50K	2,915	0.467	0.047	0.335	0.727
	3K	2,915	0.469	0.042	0.358	0.719
	1K	2,915	0.469	0.045	0.334	0.710
	400	2,915	0.469	0.056	0.320	0.743
	151	2,915	0.469	0.087	0.187	0.826
MO	50K	2,708	0.466	0.045	0.344	0.736
	3K	2,708	0.468	0.039	0.363	0.727
	1K	2,708	0.472	0.046	0.322	0.731
	400	2,708	0.474	0.057	0.297	0.728
	151	2,708	0.462	0.089	0.168	0.798
PHS	50K	186,716	0.236	0.053	-0.072	0.575
	3K	186,716	0.235	0.054	-0.092	0.566
	1K	186,716	0.232	0.058	-0.109	0.585
	400	186,716	0.231	0.070	-0.205	0.622
	151	186,716	0.241	0.103	-0.214	0.765
MHS	50K	1,560	0.251	0.091	0.082	0.621
	3K	1,560	0.252	0.090	0.071	0.637
	1K	1,560	0.248	0.096	0.045	0.633
	400	1,560	0.250	0.112	-0.014	0.697
	151	1,560	0.254	0.122	-0.154	0.760
GMGO	50K	5,327	0.239	0.059	0.063	0.539
	3K	5,327	0.234	0.058	0.048	0.499
	1K	5,327	0.233	0.062	0.014	0.484
	400	5,327	0.231	0.075	-0.008	0.523
	151	5,327	0.233	0.106	-0.100	0.634

[A]SS = self-self; FS = full sibs; FO = father – offspring; MO = mother – offspring; PHS = paternal half-sibs; MHS = maternal half-sibs; GMGO = grand-mother – grand-offspring.

Notably, the distribution of estimated genomic relationship for full-sibs was not centred at the expected value of 0.5 and instead averaged ~0.47 for all SNP panels considered (Figure 1, left panel). This same anomaly was reported by Lourenco *et al* (2015) and was attributed to both genotype and pedigree errors. Indeed, with the possible exception of genomic relationships estimated for self-self and for maternal half-sibs which was centred at the expected value of 1.0 and 0.25, respectively, all other relationships were centred at a value slightly lower than the expectation. Further research is needed to ascertain whether errors in pedigree and/or genotypes are responsible for this anomaly.

Table 2 presents the correlation between genomic relationships estimated using the 50K SNP chip and the four low-density SNP panels. On average, this correlation decreased from 0.864 when using the 3K SNP panel to 0.465 when using the 151 SNP panel. However, the decrease was not linear, with the smallest being by 10.7% from the 3K to the 1K panel (0.864 to 0.771), and the largest by 27.1% from 400 to 151 SNP panels (0.638 to 0.465).

Table 2. Correlation between genomic relationships estimated using the high-density 50K SNP chip and four low-density SNP panels for seven types of pedigree-based kinships

Panel	Type of Kinship[A]							Average
	SS	FS	FO	MO	PHS	MHS	GMGO	
3K	0.752	0.892	0.834	0.807	0.889	0.966	0.905	0.864
1K	0.607	0.808	0.729	0.699	0.797	0.933	0.828	0.771
400	0.455	0.682	0.595	0.501	0.654	0.869	0.713	0.638
151	0.292	0.484	0.411	0.341	0.470	0.732	0.525	0.465
Average	0.526	0.716	0.642	0.587	0.702	0.875	0.743	

[A]SS = self-self; FS = full sibs; FO = father – offspring; MO = mother – offspring; PHS = paternal half-sibs; MHS = maternal half-sibs; GMGO = grand-mother – grand-offspring.

Averaged across the four low-density panels, self-self relationships (from diagonal elements of the GRM) were the least correlated (r = 0.526) with the ones obtained with the 50K panel, followed by mother-offspring (r = 0.587) and father-offspring (r = 0.642). The highest average correlation was observed for maternal half-sib combinations (r = 0.875).

CONCLUSIONS

In contrast to other livestock species, broiler chicken have large full-sib families implying a large benefit in adopting genomic evaluation compared to pedigree-based evaluation. However, this benefit relies on accurate estimation of relationships among individuals. This accuracy is affected when using low-density panels as a cost-effective alternative to genomic evaluation with a 50K panel. We conclude that a panel of 1,000 SNP can be used to reliably estimate relationships. However, further research is needed to ascertain the potential impact on the breeding goal of a selection line when the SNP in a low-density panels have been selected according to a fitness function that includes the association of SNP to traits in the breeding objective.

REFERENCES

Badke YM, Bates RO, Ernst CW, Fix J, and Steibel JP (2014) *G3 (Bethesda)* **4**: 623.
Chapuis H, Pincent C, and Colleau JJ. (2016) *J. Anim. Breed. Genet*. **133**: 3.
Judge MM, Kelleher MM, Kearney JF, Sleator RD, and Berry DP (2016a) *Animal* **24**: 1.
Kirkpatrick S, Gelatt CD Jr, and Vecchi MP. (1983) *Science* **220**: 671.
Lourenco DA, Fragomeni BO, Tsuruta S, Aguilar I, Zumbach B, et al. (2015) *Genet. Sel. Evol*. **47**: 56.
Ogawa S, Matsuda H, Taniguchi Y, Watanabe T, Takasuga A, et al. (2016) *Anim. Sci. J*. **87**: 3.
Schierenbeck S, Pimentel EC, Tietze M, Körte J, Reents R, et al. (2011) *J. Dairy Sci*. **94**: 6143.
Strucken EM, Lee SH, Lee HK, Song KD, Gibson JP, et al. (2016) *J. Anim. Breed. Genet.* **133**: 13.
VanRaden PM. (2008) *J. Dairy Sci*. **91**: 4414.
Ventura RV, Miller SP, Dodds KG, Auvray B, Lee M, et al. (2016) *Genet. Sel. Evol*. **48**: 71.
Wang C, Habier D, Peiris BL, Wolc A, Kranis A, et al. (2013) *Poult. Sci.* **92**: 1712.