Estimation of heritability of body weight in broilers using pedigree and dense genome-wide SNP data

A. Kominakis[1], E. Tarsani[1], G. Maniatis[2], V. Olori[2] & A. Kranis[2,3]

[1]Department of Animal Science, Agricultural University of Athens, Iera Odos, 11855, Athens, Greece (acom@aua.gr) (Corresponding Author)
[2]Aviagen, EH28 8SZ, Midlothian, United Kingdom
[3] The Roslin Institute, University of Edinburgh, EH25 9RG, Midlothian, United Kingdom

Summary

We used pedigree and dense genome-wide SNP data on 6,598 broilers to obtain heritability estimates for body weight at 35 days of age using two types of genomic relationship matrices, seven minor allele frequency classes and four proportions of numbers of SNPs. We also attempted to partition genomic additive variance into individual autosomal contributions. The pedigree heritability estimate was 0.327 while the maximum genomic heritability (0.29) was attained when the kinship matrices were constructed on common variants and genetic relationships were estimated on identity by state alleles sharing metrics. Genomic heritability was independent on the number of SNPs used and only 10% of randomly selected SNPs were enough to capture the (co)variance of animals' additive effects. A strong linear relationship between the proportion of variance explained by each chromosome and chromosome length was observed with longer chromosomes explaining larger proportion(s) of genetic variance.

Introduction

Narrow sense-heritability (h^2) has traditionally been estimated with the usage of relatedness via known pedigree ($\hat{h}^2_{ped}$) and the additive (numerator) relationship matrix (A) of animals. Recently, estimation of h^2 can also be implemented via the genomic relationship matrix (GRM, van Raden, 2008) of animals derived from dense genome-wide SNP data (Visscher et al. 2008). This estimation of h^2 is termed the 'pseudo-heritability' ($\hat{h}^2_{ps}$) (Kang et al. 2010). In the present study, we used pedigree and dense genome-wide SNP data on 6,598 broilers with records of body weight at 35 days of age (BW35) with the aim: i) to compare heritability estimates based on pedigree and genomic data, ii) to compare estimates of $\hat{h}^2_{ps}$ based on the GRM and the Identity By State (IBS) distance matrix, iii) to assess the relationship between MAF (Minor Allele Frequency) and $\hat{h}^2_{ps}$ and iv) to partition genomic additive variance into autosomal contributions.

Material and methods

Data

A total of n=6,598 broilers (n=3,678 males and n=2,920 females) from a purebred commercial broiler line with records on BW35 were made available by Aviagen Ltd. Animals were genotyped using an Affymetrix® Axiom® high-density genotyping array. After applying quality control filters (only autosomal polymorphic SNPS with call rate >0.99 and

autosomal heterozygosity outside the 1.5 inter-quartile range; Turner et al. 2011) the final number of the SNPs used in this study was 450,885. Data filtering was performed using the SNP & Variation Suite v8.7.2 software (Golden Helix: http://www.goldenhelix.com). Phenotypic records for BW35 ranged from 1,130 to 2,630 g with an average of 1840.2 g (SD=194 g). The pedigree data, which is typically available for broilers, consisted of 730 full-sibs families with an average of 8.7 full-sibs per family (ranging from 2 to 48) and an average relationship of 0.28765 (SD=0.110) (min=0.0625, max=0.50). A maximum number of 2.0 discrete generation equivalents (Boichard et al. 1997) were also calculated for this data set. Pedigree analysis was carried out using the CFC software tool (Sargolzaei et al. 2006).

Estimation of pseudo-heritability

Variance components were estimated using a multi-locus mixed-model stepwise regression with forward inclusion and backward elimination (Segura et al. 2012) in the following cases: (i) two methods for the construction of the animals' kinship relationships: one based on the GRM and a second method based on the IBS pairwise distances of samples defined as d_{IBS}=[(number of markers with two IBS alleles)+0.5*(number of markers with one IBS allele)]/(number of non-missing markers) (Purcell et al. 2007), ii) seven MAF classes i.e. <0.01, 0.01-0.05, 0.05-0.10, 0.10-0.20, 0.20-0.30, 0.30-0.40 and 0.40-0.50 and iii) four proportions of numbers of SNPs used to construct the GRM.

Each time, the random additive genetic effect of individual animals was incorporated in the mixed model with a covariance structure based on the GRM or the IBS distance matrices, the MAF classes and the proportion of SNPs used. The fixed effects part of the model included hatch week (36 classes), mating group (17 classes) and sex (2 classes). We further partitioned genomic heritability attributable to each autosome by fitting all autosomes in a full model and then drop them one at a time for a reduced model. This strategy allowed us to assess the genetic variance associated with particular chromosome(s) while accounting for background polygenic effects on other chromosomes (Rowe et al. 2014). The optimal step(s) during forward inclusion and backward elimination across the various cases was selected using the extended Bayesian Information Criterion (eBIC) (Chen and Chen, 2008). All the above analyses were performed with the SNP & Variation Suite v8.7.2 software. We also estimated the heritability for the trait using the A matrix ($\hat{h}^2_{ped}$) via the ASREML software (Gilmour et al. 2009).

Results

Table 1 shows the variance components as well as the $\hat{h}^2_{ps}$ estimates across the various MAF classes and the two scenarios for constructing the animals' kinship matrix. To make results independent from the number of markers used, the same number of SNPs (n=42,044) was aimed for all MAF classes. Note though, that due to the typical U-shape distribution of SNPs along the MAF classes, a smaller number of markers were used (n=35,509) in the case of 0.20<MAF<0.30. The $\hat{h}^2_{ped}$ estimate was 0.327 (SE=0.032). The $\hat{h}^2_{ps}$ progressively increased as MAF increased and reached a maximum value (0.29) at 0.20<MAF<0.30 when the IBS distance matrix was implemented. Generally, implementation of the IBS distance matrices

always resulted in higher $\hat{h}^2_{ps}$ when contrasted to GRM based estimates (Table 1).

Table 2 shows $\hat{h}^2_{ps}$ when randomly selected subsets of 451 K SNPs were used to construct the GRM. As Table 2 displays, $\hat{h}^2_{ps}$ did not depend on the number of SNPs used and using only 10% of randomly selected SNPs, regardless of their MAF, was enough to capture the (co)variance of animals' additive effects. Figure 1 displays the proportion of genetic variance explained (PVE, %) by each autosome. A strong linear relationship between the PVE by each chromosome and chromosome length (L, in Mb) (PVE=0.75+0.0766(±0.0012)*L, r^2=0.983, p<0.001) was detected denoting that longer chromosomes explained a larger proportion of genetic variance and therefore are likely to harbor more quantitative trait loci.

Discussion

Heritability estimates based on genomic data are expected to be more accurate when contrasted to pedigree estimates, since genomic data capture more accurately the Mendelian sampling and thus should better reflect the genetic relationships of animals (Vinkhuyzen et al. 2013). In the present study, the highest $\hat{h}^2_{ps}$ was 0.29 while the pedigree estimate was 0.33. Vinkhuyzen et al. (2013) reported heritability estimates for various complex traits in humans, which have a very high effective population size, and concluded that the use of genomic relationships captures only one to two thirds of the pedigree estimates. In contrast to these findings, the difference between the two estimates in the present study was small (0.04) implying a smaller effective population size and SNPs capturing much larger genetic relationships compared to human data.

We additionally observed a nonlinear pattern of $\hat{h}^2_{ps}$ across the various MAF. This partly agrees with Speed et al. (2016) who suggested that the estimate in question varies with MAF(1-MAF)$^{0.75}$. A maximal value for $\hat{h}^2_{ps}$ was obtained when the kinship matrices were constructed on common variants (0.20<MAF<0.40) since the use of rare variants results in a less accurate description of animals' relationships. Furthermore we found that $\hat{h}^2_{ps}$ can be accurately estimated when even 10% (~45K) randomly selected SNPs were used to construct the GRM. Given the total length of chicken genome (~1000 Mb) and the corresponding average SNP density when using 45K SNPs across the genome, the amount of LD between markers is high (Andreescu et al., 2007) in contrast with human populations that have very large effective population size. The result also agrees with Yang et al. (2010) who suggested that the variance explained by genome-wide SNPs remains insensitive to the proportion of SNPs used to estimate adjusted genetic relationships.

The partitioning of additive genetic variance to contribution by individual autosomes was consistent with a highly polygenic architecture with a strong linear relationship between the estimate of variance explained by each autosome and chromosome length. A similar observation was derived in human populations when a typical polygenic trait such as height was studied using a large genomic dataset (Yang et al. 2010).

References

Andreescu C, Avendano S, Brown SR, Hassen A, Lamont SJ and Dekkers JCM 2008.

Linkage Disequilibrium in Related Breeding Lines of Chickens. Genetics 177: 2161-2169.
Boichard D., Maignel L., Verrier E. (1997): The value of using probabilities of gene origin to measure genetic variability in a population. Genetics Selection Evolution, 29, 5–23.
Chen JH and Chen ZH, 2008. Extended Bayesian information criteria for model selection with large model spaces. Biometrika;95:759–771.
Gilmour et al. 2009. ASReml User Guide Release 3.0. VSN International Ltd.. Hemel Hempstead, UK.
Kang et al. 2010. Variance component model to account for sample structure in genome-wide association studies. Nature genetics.;42(4):348-354. doi:10.1038/ng.548.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559-75. Epub 2007 Jul 25.

Rowe et al. 2014. Analysis of the genetics of boar taint reveals both single SNPs and regional effects. Rowe et al. BMC Genomics 2014, 15:424.
Sargolzaei M., Iwaisaki H., Colleau J.J. (2006): CFC: a tool for monitoring genetic diversity. In: Proc. 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil, 27–28.
Segura et al. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet. 44:825–830. doi: 10.1038/ng.2314.

Speed et al. 2012. Improved Heritability Estimation from Genome-wide SNPs. Am. Journal of Human Genetics, http://dx.doi.org/10.1016/j.ajhg.2012.10.01.
Speed et al. 2016. Re-evaluation of SNP heritability in complex human traits. doi: http://dx.doi.org/10.1101/074310.
Turner et al. 2011. Quality Control Procedures for Genome Wide Association Studies. Curr Protoc Hum Genet. 2011; Unit1.19, doi: 10.1002/0471142905.hg0119s68.
vanRaden, PM, 2008. Efficient methods to compute genomic predictions. J Dairy Sci. 2008 Nov;91(11):4414-23. doi: 10.3168/jds.2007-0980.
Vinkhuyzen et al. 2013. Estimation and Partitioning of Heritability in Human Populations using Whole Genome Analysis Methods. Annu Rev Genet. 2013; 47: 75–95. doi: 10.1146/annurev-genet-111212-133258.
Visscher et al. 2008. Heritability in the genomics era —concepts and misconceptions. Nat. Rev. Genet. 9:255–266, doi:10.1038/nrg2322.
Visscher, P.M. et al. (2008) Heritability in the genomics era–concepts and misconceptions. Nat. Rev. Genet., 9, 255–266.
Yang et al. 2010. Common SNPs explain a large proportion of heritability for human height. Nature genetics. 42(7):565-569. doi:10.1038/ng.608.

Table 1. Pseudo-heritability ($\hat{h}^2_{ps}$) of BW35 at seven MAF classes and two kinship matrices

		MAF						
Matrix		**<0.01**	**0.01-0.05**	**0.05-0.10**	**0.10-0.20**	**0.20-0.30**	**0.30-0.40**	**0.40-0.50**
GRM	$\hat{\sigma}^2_A$	754.9	2086.5	4241.2	3920.9	4033.6	3731.7	3168.0
	$\hat{\sigma}^2_e$	17421.2	15168.7	14091.6	13304.6	13451.5	13593.3	14175.9
	$\hat{h}^2_{ps}$	0.042	0.121	0.231	0.228	0.230	0.215	0.183
IBS	$\hat{\sigma}^2_A$	829.6	2280.7	5006.9	4445.0	5045.6	4932.1	4241.3
	$\hat{\sigma}^2_e$	17470.4	15016.6	13728.1	12654.8	12445.0	12633.5	13343.7
	$\hat{h}^2_{ps}$	0.045	0.132	0.267	0.260	0.288	0.281	0.241

σ^2_A: additive genetic variance, σ^2_e: error variance, $h^2_{ps} = \hat{\sigma}^2_A / (\hat{\sigma}^2_A + \hat{\sigma}^2_e)$

Table 2. Pseudo-heritability ($\hat{h}^2_{ps}$) of BW35 in relation to number of SNPs used to construct the GRM

Number of SNPs (proportion %)	$\hat{\sigma}^2_A$	$\hat{\sigma}^2_e$	$\hat{h}^2_{ps}$
450,885 (100)	4951.4	12441.0	0.285
225,442 (50)	4992.3	12398.1	0.287
45,090 (10)	4987.6	12713.1	0.282
22,545 (5)	4674.0	13120.8	0.263

σ^2_A: additive genetic variance, σ^2_e: error variance, $h^2_{ps} = \hat{\sigma}^2_A / (\hat{\sigma}^2_A + \hat{\sigma}^2_e)$

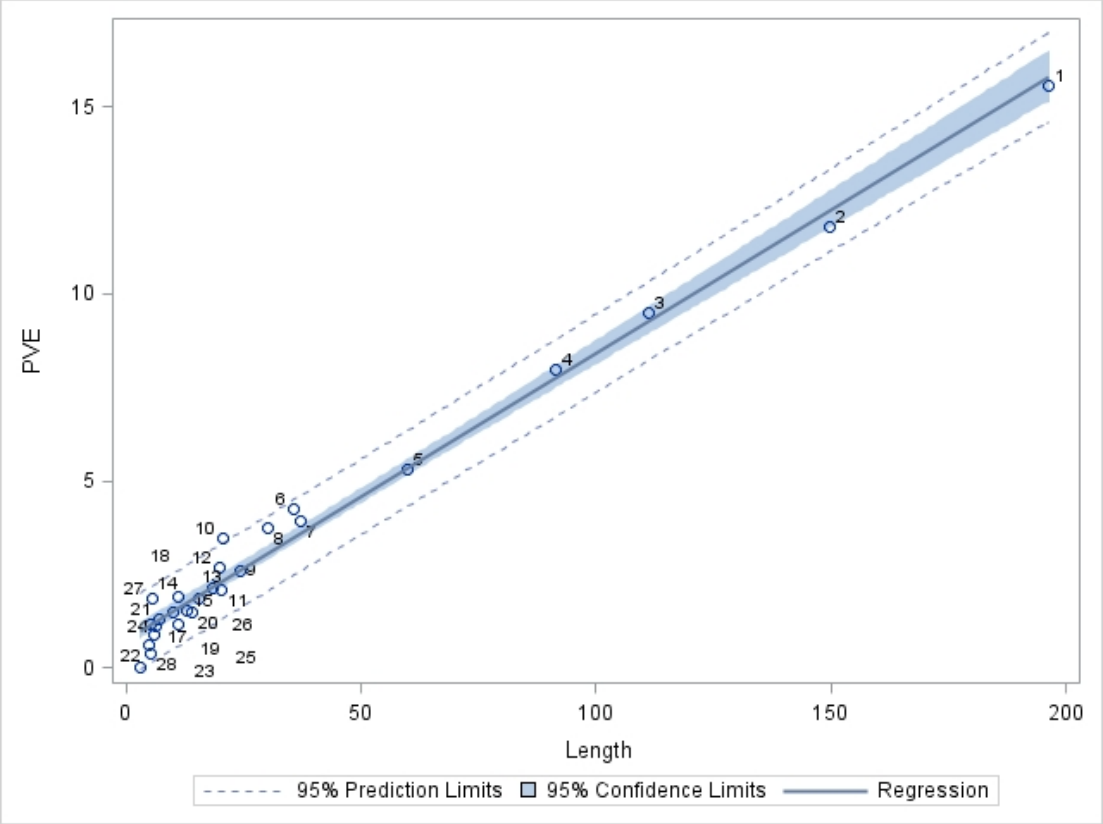

Figure 1. Relationship between Proportion of Variance Explained (PVE,%) by each autosome and chromosome length (in Mb). Numbers 1 to 28 denote chromosomes.