https://doi.org/10.17221/83/2017-CJAS

Genomic Evaluation and Variance Component Estimation of Additive and Dominance Effects Using Single Nucleotide Polymorphism Markers in Heterogeneous Stock Mice

Morteza Mahdavi[1,2], Gholam Reza Dashab[1], Mehdi Vafaye Valleh[1], Mohammad Rokouei[1,3], Mehdi Sargolzaei[4,5]*

[1]*Department of Animal Science, University of Zabol, Zabol, Iran*

[2]*Arak Branch, Razi Vaccine and Serum Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Arak, Iran*

[3]*Department of Bioinformatics, University of Zabol, Zabol, Iran*

[4]*Department of Pathobiology, University of Guelph, Guelph, Canada*

[5]*HiggsGene Solutions Inc., Guelph, Canada*

**Corresponding author: dashab@uoz.ac.ir*

ABSTRACT

Mahdavi M., Dashab G.R., Vafaye Valleh M., Rokouei M., Sargolzaei M.: (2018): **Genomic evaluation and variance component estimation of additive and dominance effects using single nucleotide polymorphism markers in heterogeneous stock mice**. Czech J. Anim. Sci., 63, 492–506.

Exploration of genetic variance has mostly been limited to additive effects estimated using pedigree data and non-additive effects have been ignored. This study aimed to evaluate the performance of single nucleotide polymorphisms (SNPs) marker models in the mixed and orthogonal framework including both additive and non-additive effects for estimating variances and genomic prediction in four diabetes-related traits in heterogeneous stock mice. Models have performed differently in detecting SNPs affecting traits. Dominance variances explained over 14.7 and 3.8% of genetic and phenotype variance in a Genomic prediction and variance component estimation method (GVCBLUP) framework. Reliabilities of additive Genomic best linear unbiased prediction model (GBLUP) in different traits ranged from 44.8 to 66.6%, for GVCBLUPs framework including both additive and dominance effects (MAD), and 46.1 to 69% for the model including additive effect (MA). Dominance GBLUP reliabilities ranged from 6 to 26.4% for MAD and from 22.5 to 50.5% in the model including dominance (MD). MA and MD had higher reliability for additive and dominance GBLUPs compared to MAD. Reliabilities of GBLUPs in MAD and MA for all traits were not significant except for growth slope ($P < 0.01$). In orthogonal framework models, epistasis variances accounted for a greater proportion (87.3, 89.1, 95.5, and 77.2%) of genetic variation for end weight, growth slope, body mass index, and body length, respectively. Heritability in a broad sense was estimated at 1.12, 1.67, 3.64, and 2.0%, in which non-additive heritability had a significant contribution. Genetic variances explained by dominance using GVCBLUPs were 16.8, 29.4, 14.6, and 14.9% for the traits. Generally, the non-additive models had a lower value of deviance information criterion (DIC) and performed better in estimating the variance component. Comparing the estimated variance by orthogonal framework models confirmed the results previously estimated by GVCBLUPs, with the difference that the estimates were shrinking. Following significant SNPs affecting diabetes-related traits by post-genome-wide studies could reveal unknown aspects and contribute to genetic control of the disease.

Keywords: orthogonal; genomic; SNP; model; mouse; genetic effect

Supported by the Wellcome Trust.

https://doi.org/10.17221/83/2017-CJAS

Traditionally, the exploration of genetic variance in humans, plants, and livestock species has mostly been limited to the use of additive effects estimated using pedigree data. In this context, the impact of genetics in complex traits has been quantified as heritability, e.g., the proportion of the total phenotypic variance explained by additive genetic variance. However, the estimation of heritability via the use of additive models does not only capture additive gene action but can potentially also capture part of the dominance effects and epistasis interactions (Hill et al. 2008). Therefore, the proportion of phenotypic variation that is explained by all genetic effects, and how much of the total genetic variation is actually due to additive effects, has still been unclear in modern genetics (Vinkhuyzen et al. 2013).

Until recently, due to the lack of genomic data, estimation of dominance deviations using phenotypic and pedigree data has been subjected to some limitation such as high computational demand and unavailability of large data with sufficient proportion of informative individuals (full-sibs) in data set. Several studies have estimated non-additive variances in livestock using traditional pedigree information (Palucci et al. 2007) and reported a small but significant non-additive variance. However, it is difficult to estimate non-additive variance because firstly it is often, at least partially, confounded with other effects such as common environment or maternal effects. Consequently, estimates of non-additive variance may be biased upwards. Secondly, there is a lack of informative pedigrees, typically with large full-sib families, which are needed for accurate estimates of dominance effects. Thirdly, at the individual animal level, dominance is hardly used in animal breeding, although it contains a relevant part of genetic variation. The reasons are the heavy computational demand of large-scale genetic evaluations for dominance, the relatively low accuracy of resulting estimates of dominance effects, and the complexity of planning and computing the outcome of planned mating. Some estimations of dominance variance in dairy cattle that are based on pedigree data range from 7.3 to 49.8% of the total genetic variance for conformation traits (Tempelman and Burnside 1990) and from 3.4 to 42.9% for milk production traits (Van Tassell et al. 2000).

In view of this, it is not surprising that most genetic evaluation systems use an additive model and ignore non-additive effects, especially considering that their aim is to estimate breeding values or additive genetic values. In addition, Hill (2008) argued that even if gene effects are non-additive, most of the genetic variance is still expected to be additive variance. Despite this, dominance has still been of theoretical and practical interest, because it is heavily used in crosses of animal breeds and plant lines (e.g., in pigs, poultry, or corn). In principle, assortative mating or mate allocation can promote the field performances of livestock and crops.

In livestock populations, one of the main reasons why dominance or higher order interaction terms have not been considered in genetic evaluations is the lack of pedigree or that the pedigree relationships are not informative enough. The recent advent of dense single nucleotide polymorphism (SNP) panels and genomic selection methods has lightened interest in the prediction of non-additive genetic effects. Most epistatic models only focus on additive-by-additive epistatic interactions (Jiang and Reif 2015) while dominance-by-dominance and dominance-by-additive interactions may play a major role in heterosis. A non-orthogonal partition of variance component may be suggested a large confounding between the estimated additive and non-additive variance and a wrong conclusion, especially in crossbreeding systems. This is probably because their additive and dominance covariates are highly correlated. In fact, the availability of SNP genotypes especially in the absence of reliable pedigree data e.g. in sheep and salmon and distinctive method and model represent a new opportunity to estimate non-additive effects at individual loci and to estimate non-additive variances.

Large-scale genome-wide association studies (GWAS) have identified thousands of SNPs associated with hundreds of traits and diseases. However, only a fraction of the trait variance can be explained by the SNPs reported by GWAS. The proportion of the total additive genetic variance explained by the quantitative trait loci (QTL) detected in the GWAS was always between 10 and 40% and even lower if expressed as a fraction of the phenotypic variance (Armando et al. 2015). This is in broad agreement with the percentages of variation explained by significant SNPs in a range of studies (Wood et al. 2014) with very large sample sizes. With the availability of SNP genotypes,

https://doi.org/10.17221/83/2017-CJAS

non-additive effects at a marker locus or between loci can be readily determined, dominance effects of markers can be estimated (Wellmann and Bennewitz 2012) and computing the expected outcome of planned mating based on SNP genotypes is straightforward. Furthermore, covariance matrices of genomic non-additive effects among individuals can be calculated, similar to matrices of genomic additive relationships, which are widely used in genomic selection, such that these effects can be estimated in the Genomic best linear unbiased prediction model (GBLUP) (Vitezica et al. 2017).

The aim of this study is to estimate genetic effects and variance components using SNP markers, including non-additive effects, and evaluate the reliability of prediction by considering additive, non-additive, and both effects in the model using methods listed and real data, heterogeneous stock (HS) mice data, build 37, to illustrate the principles.

MATERIAL AND METHODS

Animals. The data (genotype and phenotype file) from eight HS mice founders, which had passed 50 generations of pseudorandom breeding, were used for genomic prediction and variance component estimation of additive and dominance effects using SNP markers. In order to know the structure of HS mice population, a genealogy investigation was performed. The population consisted of 172 dams and 172 sires and 1994 progeny with 0.0041 average numerator relationships. Pedigree extended over 2284 individuals. This genealogy was organised into 165 full-sib families with 11.7 offspring on average.

Genotype. We selected SNPs across the genome that distinguish between the eight HS founders. We used datasets to select SNPs that are validated and polymorphic in at least some of the HS founders. Data are freely available at http://mtweb.cs.ucl.ac.uk/mus/www/mouse/HS/index.shtml. This population is valuable for testing genome-wide studies because, due to the high number of markers, it is expected that many (about three of every five) QTL loci will be in complete linkage disequilibrium (LD) with marker loci (Mott et al. 2000). Only animals with available phenotype and genotype were used for data analysis. Our data set was composed of 1940 individuals with 12 226 polymorphic loci (SNPs). Genotypes with minor allele frequency lower than 0.001 were excluded. These missing values should have a negligible effect on the analysis.

A script code in R package was used in order to convert original genotype format to a format readable for software. The coding for SNP genotypes was done as follows: always A1A1 = 0, A1A2 = 1, and A2A2 = 2, where '0' and '2' denote the two homozygous genotypes and '1' denotes the heterozygous genotype. Any other number denotes a missing SNP genotype and results in zero in matrix construction. For the dominant component coding for SNP genotypes A1A1, A1A2, and A2A2 were set 0, 1, and 0, respectively.

Traits and phenotypes. The phenotype file contains four quantitative diabetes-related traits (weight at 10 weeks, growth slope, body mass index (BMI), and body length) which were used as response variables. All probable covariates and fix effects according to a model menu presented in data package were previously tested for affecting traits and significant covariates and fixed effects for each trait were included in the main models. These covariates were sex for weight and growth slope, and body weight, season, month and day for BMI and body length.

Statistical method and models. Limited methods on genomic prediction and variance component estimation of dominance were available. Genomic best linear unbiased prediction (GBLUP) and various Bayesian methods are available for genomic prediction; GBLUP generally had good performance in real data (de los Campos et al. 2013). Restricted maximum likelihood estimation (REML) and Gibbs sampling in the case of small sample size, which has been a widely accepted method for estimating variance components, were used in this study. Three methods, with an expansion of the additive model to dominance and epistasis in four seniors, were applied. Seniors were as follows:

(MA): $Y = \mu + XB + g_A + e$

(MD): $Y = \mu + XB + g_D + e$

(MAD): $Y = \mu + XB + g_A + g_D + e$

(MADE): $Y = \mu + XB + g_A + g_D + g_E + e$

where:

μ = overall mean

XB = fixed effects

g_A, g_D, g_E = additive, dominance, epistasis effects, respectively

e = error

https://doi.org/10.17221/83/2017-CJAS

MA includes genomic information in order to estimate additive effects (g_A) on marker genotypes, MD estimates just dominance effects (g_D), MAD considers both additive (g_A) and dominance (g_D), and MADE adds epistasis effects (g_E) to the MAD scenario. A simple Genomic prediction and variance component estimation method (GVCBLUP), an Expansion of Natural and Orthogonal Interactions (NOIA) approach, and a method including epistasis effects (PEPIS) are three methods by which the scenarios were analysed. PEPIS was carried out using a web server-based tool, the Pipeline for estimating epistatic genetic effects, for analysing the epistatic effects (Zhang et al. 2016). The method of including epistasis effects, PEPIS, is performed by dividing the analysis into two parts: the first part, calculating six kinship matrix (Ka, Kd, Kaa, Kad, Kda, Kdd), and a second part, calculating six component ratio estimations and further genome scanning for main and epistasis genetic effects.

All analyses in the simple Genomic prediction and variance component estimation framework were carried out using the GVCBLUP computer package for genomic prediction and variance component estimation of additive and dominance effects using SNP markers (Wang et al. 2014) and GIBBSF90 software. GVCBLUP method models are re-parameterised models of the original quantitative genetics model using a re-parameterised μ, under the assumption of equal allele frequencies. Re-parameterised models are the basis for the SNP coding of 0-1-2 for additive effect and 0-1-0 coding for dominance effect (Yang et al. 2014). GVCBLUP was set to use AI-REML for estimating variance components. If AI-REML produced any negative estimate of variance components, the program returned to EM-type REML automatically. The AI-REML algorithm is fast but is not as reliable as the EM type. The number of iterations was set at 1 000 000. Estimated additive and dominance effects of SNP markers were then printed in an output file to be directly used as the input file for graphical viewing of marker effects by SNPEVEG (Wang et al. 2012), including Manhattan plots and figures by chromosome.

Because genetic effect covariates in practice are highly correlated, variance estimates by GVCBLUP method may be inflated in both joint model (MAD) and single component model (MA, MD). To deal with these issues, we used two additional methods. First, we did an Expansion of NOIA orthogonal approach that builds incidence matrices based on genotypic frequencies in order to include genome-wide epistasis in genomic evaluation according to the method of Vitezica et al. (2017). Later the analysis was done with only fits epistasis effects using the method of Zhang (2016). The orthogonal estimation method is a development of a general procedure to estimate "genomic" relationship matrices for interaction terms of any order expanding the NOIA approach (Alvarez-Castro and Carlborg 2007). To do this and because of a small number of samples, the (co)variance components and genetic parameters were estimated through the Gibbs sampler algorithm by using the GIBBSF90 software. A total of 1 000 000 samples were generated, assuming a burn-in period of 100 000 iterations. The convergence was assessed by Geweke test using POSTGIBBSF90 software (Misztal et al. 2002). Orthogonal group models were set considering just additive effect (NOIA: MA), both additive and dominance effects (NOIA: MAD), and epistasis effect (NOIA: MADE).

Statistical significance of the associations between SNPs and traits was tested based on Bonferroni-corrected *P*-values. This method treats individual tests as independent and thus is very conservative. Aiming for an overall false positive rate of 0.05 and considering 12 500 independent tests, the point-wise *P*-value should be between 0.0001 and 0.0004. In this article, individual SNPs with a *P*-value of 0.0001 or lower were considered statistically significant.

RESULTS AND DISCUSSION

Estimate of variance components. Estimates of variance explained and their standard errors for different traits and methods are shown in Table 1. After estimation of SNP effects, additive, dominance, epistasis, and residual variances were estimated using the complete set of SNPs. Genetic variances were estimated using the additive genomic relationship (MA), dominance genomic relationship (MD), both additive and dominance genomic relationship (MAD) and all kinship matrices, Ka, Kd, Kaa, Kad, Kda, Kdd (MADE) for each trait. Table 2 shows more details of the orthogonal estimation of variance explained according to their importance in the joint model

https://doi.org/10.17221/83/2017-CJAS

or single component model. Estimates of the partitioning of the genetic variance clearly differ; for instance, under the GVCBLUP MA and MAD, the proportions of genetic variance explained by additive genetic variance were 100 and 83.2% while these values were estimated at 100, 67, and 3.9% by NOIA: MA, NOIA: MAD, NOIA: MADE models, respectively. As was expected, the standard error (SE) of estimates was greater in GVCBLUP than in orthogonal estimation. This may be due to a desirable property in modelling and amount of statistical information to estimate variances which is greater with orthogonal than with non-orthogonal models.

Generally, the models perform better in estimating variances when non-additive effects are included. As shown in Table 2, the models with non-additive effects had a lower value of deviance information criterion (DIC) than models with just additive effects, however, estimating variance components by GVCBLUP yields biased estimates, inflates the total genetic variance. Unlike GVCBLUP, NOIA: MA, MAD, and MADE have an orthogonal property which is very important and useful. As shown in Table 1, in GVCBLUP estimations due to the absence of epistatic effects in models, the additive variance has the greatest contribution to the formation of genetic variance which seems to be confounded with additive × additive part of epistatic variance. In PEPIS method, with the introduction of epistatic effects in the models, the highest contribution of genetic variance is made especially by epistatic variance. But in NOIA method, with the independent estimation of genetic effects, the results have been moderated and the genetic variance explained by epistatic effects has been decreased.

The results of the two methods, GVCBLUP (MA, MAD) and NOIA (NOIA: MA, NOIA: MAD, NOIA: MADE) models, each one in their own group, were compared. The model MAD estimates both the additive and dominance variance but the output of the MA and MD models is only additive or dominance variance, respectively. Two additive variances estimated from MAD, MA and two dominance variances estimated from MAD, MD in each trait were compared. Estimates of additive variance for end weight and BMI obtained from MAD were similar to those obtained from MA but the variances for growth slope and body length were estimated differently ($P < 0.01$). Estimates of dominance variance obtained using the MAD model for all traits were completely different from those using the MD model ($P < 0.01$). Residual variances from three GVCBLUP models (MAD, MA, and MD) were similar except for growth slope

Table 1. Estimates of variance explained and their standard errors for different traits and methods[1]

Trait	Model	Va/Vg	SE	Vd/Vg	SE	Ve/Vg	SE
End weight	GVCBLUP-MAD	0.832	0.0229	0.168	0.0166	–	–
	PEPIS	8.76E-07	0.00066	8.76E-07	0.00066	0.999	0.022
	NOIA	0.0393	0.0009	8.77E-02	0.0027	0.8730	0.017
Growth slope	GVCBLUP-MAD	0.7056	0.032	0.294	0.026	–	–
	PEPIS	0.2787	0.28	0.0121	0.0648	0.7092	0.353
	NOIA	0.04427	0.00111	0.06412	0.00197	0.8916	0.017
Body mass index	GVCBLUP-MAD	0.854	0.0227	0.1458	0.0243	–	–
	PEPIS	0.000207	0.01	0.000207	0.01	0.999	0.022
	NOIA	0.00746	0.00014	0.0373	0.00105	0.9552	0.015
Body length	GVCBLUP-MAD	0.85	0.023	0.1494	0.0239	–	–
	PEPIS	0.000207	0.010	0.000207	0.010	0.999	0.022
	NOIA	0.161	0.00042	0.0663	0.00157	0.7726	0.015

GVCBLUP-MAD = simple genomic prediction and variance component estimation method (GVCBLUP) considering both additive and dominance effects, NOIA = Expansion of Natural and Orthogonal Interactions (NOIA) approach, PEPIS = method including epistasis effects; Va, Vd, Ve, Vg and SE are additive, dominance, epistasis, genetic variance, and related standard error, respectively

[1]GVCBLUP method considering additive and dominance effect, NOIA and PEPIS with additive, dominance, and epistasis effect were presented

https://doi.org/10.17221/83/2017-CJAS

Table 2. Variance explained estimated by the NOIA method and different scenarios

Trait	Model	Va/Vg	SE	Vd/Vg	SE	Ve/Vg	SE	DIC
End weight	MADE	0.0393	0.0009	8.77E–02	0.0027	0.8730	0.017	9434.7
	MAD	0.67076	0.19015	0.32924	0.19015	–	–	9518.04
	MA	100	–	–	–	–	–	9523.8
Growth slope	MADE	0.04427	0.00111	0.06412	0.00197	0.8916	0.017	–5978.88
	MAD	0.43885	0.17923	0.56115	0.17923	–	–	–6159.17
	MA	100	–	–	–	–	–	–6135.67
Body mass index	MADE	0.00746	0.00014	0.0373	0.00105	0.9552	0.015	–5399.13
	MAD	0.18936	0.21337	0.81064	0.21337	–	–	–5560.53
	MA	100	–	–	–	–	–	–5411.72
Body length	MADE	0.161	0.00042	0.0663	0.00157	0.7726	0.015	3143.91
	MAD	0.53269	0.2275	0.46731	0.22746	–	–	3162.46
	MA	100	–	–	–	–	–	3168.52

MADE = single nucleotide polymorphism (SNP) marker model considering additive and non-additive (dominance and epistasis) effects, MAD = SNP marker model considering additive and non-additive (dominance) effects, MA = SNP marker model considering additive effects; Va, Vd, Ve, Vg and SE are additive, dominance, epistasis, genetic variance and related standard error, respectively; DIC = deviance information criterion

($P < 0.01$). Comparison of estimated variance by the NOIA group models (Table 2) confirmed the results previously estimated by GVCBLUPs, with the difference that the estimates were shrinking. Additive variances for end weight and BMI obtained from NOIA: MADE were similar to those obtained from NOIA: MA, but these variances for growth slope and body length were estimated differently ($P < 0.01$). Generally, dominance variances estimated by GVCBLUP-MAD compared to MD were small for all traits except for growth slope in which dominance variance accounted for 29.5% and slightly less than 10%, of the genetic and phenotypic variance, respectively. Based on our results, the estimated dominance variance in proportion to additive genetic variance was about 20, 42, 17, and 17.5%, respectively, which is in the range of variances reported in several previous studies using a model with or without a pedigree-based relationship matrix. For instance, in pigs, significant contributions of non-additive genetic variance have been reported. The ratios of dominance variance to additive genetic variance ranged from 11 to 31% for reproductive and growth traits in Yorkshire pigs (Su et al. 2012). In pigs, it was reported that the proportion of non-additive variance relative to the entire QTL variance exceeded 50% in most meat quality and carcass composition traits in a porcine Duroc × Pietrain population (Grosse-Brinkhaus et al. 2010). In beef cattle, the ratio of dominance variance to additive genetic variance was larger than 50% for weaning weight in Hereford, Gelbvieh, and Charolais beef cattle (Duangjinda et al. 2001). In chicken, QTL analysis revealed that the non-additive genetic effect was more pronounced prior to 46 days of age, whereas additive genetic effect explained the major portion of the genetic variance later in life (Carlborg et al. 2004). In orthogonal models, the contribution of non-additive effects has increased in the creation of genetic variance. With NOIA: MD model non-additive goes into two parts: first, contributes to informing dominance and second, ambiguously mixes with additive variance, however with including epistasis most of the genetic variance goes to an epistatic section of variances. Hence, epistasis variances, especially additive × additive interaction, account for a greater proportion (87.3, 89.1, 95.5, and 77.2% for end weight, growth slope, BMI, and body length, respectively) of genetic variation. There is a large variation between the estimates of non-additive genetic variances in different studies, which may depend on the different genetic architecture of various traits and populations. In addition, the large variation could be caused by a large sampling error due to insufficient data information. The study methods, as shown here, may lead to a variation in estimations.

https://doi.org/10.17221/83/2017-CJAS

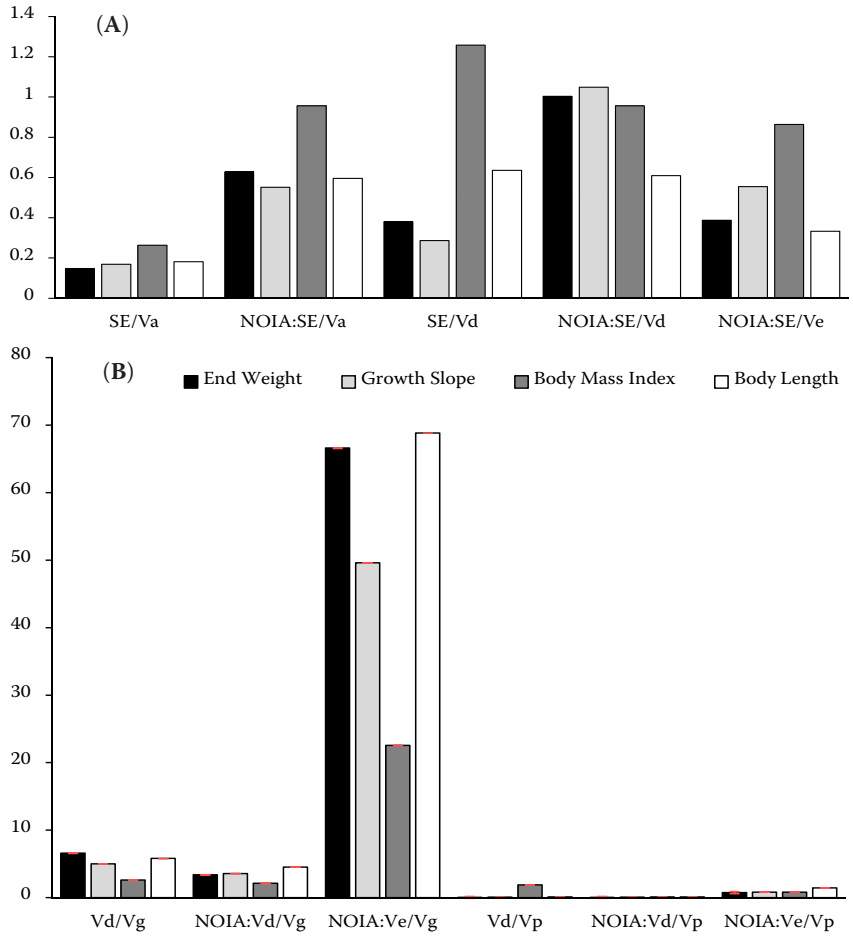

Figure 1. (**A**) Proportion of standard error to related variance estimated by conventional and orthogonal NOIA model. SE/Va, SE/Vd, and SE/Ve = proportion of related standard error to additive, dominance and epistatic variance, respectively. (**B**) Genetic and phenotypic variance explained for different traits estimated by conventional and orthogonal NOIA model. Vd/Vg, Vd/Vp, Ve/Vg, Ve/Vp indicate genetic and phenotypic variance explained by dominance and epistatic variance

As shown in Figure 1A, the standard errors of non-additive genetic variance were large. The relative standard errors for the estimated dominance variance were 1.7, 2.6, 3.5, and 4.8 times as large as the standard errors for the estimated additive genetic variance for growth slope, end weight, body length, and BMI, respectively. These relative standard errors derived from orthogonal models were 1.6, 1.9, 1.0, and 1.0 times greater if compared to GVCBLUPs, respectively. These results suggest that a large dataset is needed in order to get accurate estimates of non-additive genetic variances.

These results illustrate the difficulties in obtaining a good estimate of non-additive variance from genomic information. Despite the difficulty of accurate estimates, results show how genomic information allows one to obtain an accurate estimation of non-additive deviations. Further, in practice, use of non-additive through the implementation of mate allocations using markers is straightforward, contrary to pedigree-based methods.

Classic estimates of heritability require pedigree data, which can be costly and difficult to acquire. As genome-wide data became widely available, genome-wide complex trait analysis has been developed, which provides an SNP-based heritability estimate. In this study using SNP markers and fitting of variance components, the common heritability, the additive heritability or heritability in the narrow sense h^2_A, the dominance heritability h^2_D, and the heritability in the broad sense H^2 were estimated for traits in different models. The proportions of additive, dominance, and epista-

https://doi.org/10.17221/83/2017-CJAS

sis variance explaining genetic variance for different models and traits were also presented in Table 1. The proportions explained by dominance in GVCBLUPs ranged from 14.7% for BMI to 29.5% for growth slope. All estimated dominance variances differed from zero ($P < 0.01$). When the epistasis effects were included in the models, the proportions explaining dominance decreased, however, non-additive effects still accounted for the majority of genetic variability in most traits. The proportion of genetic and phenotype variance explained by dominance and epistasis variance in GVCBLUP and NOIA model is shown in Figure 1B. Orthogonal models show lower level of genetic and phenotypic variance explained by dominance (Vd/Vg and Vd/Vp), in comparison with GVCBLUP which confirms shrinkage estimation of variances by the orthogonal method. The method develops a general procedure to estimate "genomic" relationship matrices for interaction terms of any order expanding the NOIA approach. GVCBLUPs estimation of H^2 was higher in MAD models for all traits compared with MA and MD model, whereas just one of the additive or dominance effects has been included in the model. Estimated values of h^2_A ranged from 0.1 to 0.322 for all traits and were almost similar in both MAD and MA model with small SE. Estimates of h^2_D were higher in dominance model compared to the MAD model for all traits. These estimated heritabilities ranged from 0.035 to 0.25 and were completely different in MAD and MD models. SE of estimated h^2_D with MD model were relatively high. Estimates of the proportion of additive, dominance, and total genetic variance to phenotypic variance (e.g., h^2_D, h^2_D, H^2), which were calculated for each SNP, are presented across chromosome (in %) (Figure 2). Heritabilities presented for all traits and models show an inflated, higher level of additive heritability estimation by MA model compared to MAD model, which may be due to confounding non-additive with the additive in the absence of dominance and epistasis in the model. The heritability level in the orthogonal models was much lower than in the others due to shrinkage estimation of variances. Estimated values of H^2 were 1.12, 1.67, 3.64, and 2% for end weight, growth slope, BMI, and body length, respectively in which non-additive heritability had a significant contribution. Generally, end weight and growth slope were more heritable than BMI and body length. The comparison of heritability of end weight and growth slope revealed that end weight had a higher level of additive heritability despite the proximity of the total inheritance (H^2) indicating end weight is controlled more additively than growth slope.

Genome-scale data provide an opportunity to estimate relatedness of individuals using molecular data and then using estimated relatedness to infer heritability from the proportion of phenotypic variance explained by genotyped SNPs (Yang et al. 2014). Genetic markers can help estimate heritability in novel ways. When phenotypes are collected on a sample of individuals whose relatedness is partially or wholly unknown, genetic markers can be used to infer relatedness between pairs of individuals, because related individuals tend to share more marker alleles than unrelated individuals. The inferred relatedness can then be correlated with phenotypic similarity, and quantitative genetic parameters, including heritability, can be estimated. This method has been applied in evolutionary studies to estimate heritability for quantitative traits in fish and plants when phenotypes and SNP markers are available but pedigree information is not. The use of the pedigree data leads to an estimate of total heritability whereas the use of SNPs to construct a marker relationship matrix and estimating the genetic variance and therefore heritability explained by the SNPs. The new whole genome methods have shown that large numbers of genetic variants with small effect explain a substantial proportion of the heritability for complex traits. Valdar (2006) fitted a standard additive genetic, common environmental error, unique environmental error model to obtain estimates of the proportion of phenotypic variance attributable to additive genetic effects. The estimations of heritability for end weight, growth slope, BMI, and body length were 0.623, 0.305, 0.132, and 0.213, respectively. Estimated heritabilities in this study using GVCBLUPs for end weight, growth slope, and BMI (0.304, 0.22, 0.108) were lower than and heritability for body length (0.21) was close to those shown by Valdar (2006). The difference may be due to using just SNPs to construct a relationship matrix in order to estimate the genetic variance and therefore models captured the heritability explained by the SNPs were included in the analysis. These results indicating SNPs marks explain a substantial proportion of the heritability for complex traits.

https://doi.org/10.17221/83/2017-CJAS

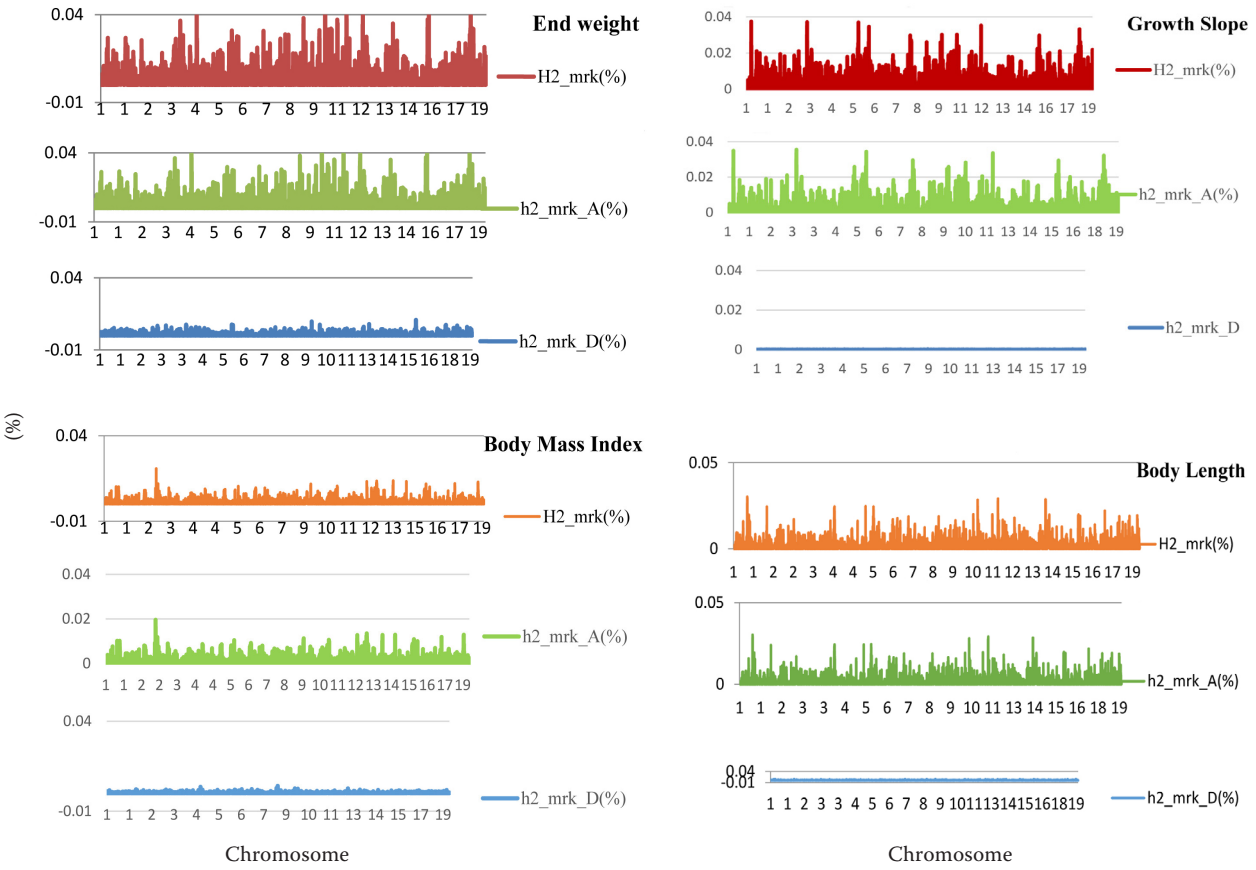

Figure 2. Percentage of heritability explained by single nucleotide polymorphisms estimated by MAD model for different traits. H2_mrk (%) = percentage of heritability in broad sense, H2_mrk_A (%) = additive heritability, h^2_mrk_D (%) = dominance heritability for different traits: end weight, growth slope, body mass index, and body length, respectively

The proportion of genetic variance explained by the SNPs depends on the structure of the data (e.g. related and unrelated individuals), the architecture of traits, linkage equilibrium (LE) or LD between SNPs, a factor that generates non-independency between loci and lack of orthogonality. So heritability estimates derived through these methods could differ in precision and possible bias. Generally, close relatives give more precision but potentially more bias, whereas distant relatives give less precision and less bias. Bias in analyses of close relatives may come from environmental variation that is confounded with additive genetic variation within families, or in the case of siblings, confounding with non-additive genetic effects (Vinkhuyzen et al. 2013). The LD affects the partition into additive, dominance, and epistatic components, such that an orthogonal partition is not possible. LD may introduce genetic covariance between different genetic effects and complicates the definition of genetic effects and the partition of the genetic variance, in particular in the presence of epistasis (Vitezica et al. 2017). Using more density SNP chips such as the whole sequence and large sample size should help find the remaining missing heritability, that is, the difference between the heritability estimates from pedigree studies and the heritability estimated from SNPs. Sequence data could be more powerful in traits where causal variants are rare.

Genome-wide association studies (GWAS) fitting additive and dominance. The Manhattan plot of the –log10 (*P*-values) of SNP dominance and additive effects estimated by MAD GVCBLUP for all traits are in Figure 3. The number of SNPs showing only a significant ($P < 10^{-4}$) additive or dominance effect and those showing both additive and dominance effects simultaneously in the training population for each trait and model are given in Table 3. For instance, for end

https://doi.org/10.17221/83/2017-CJAS

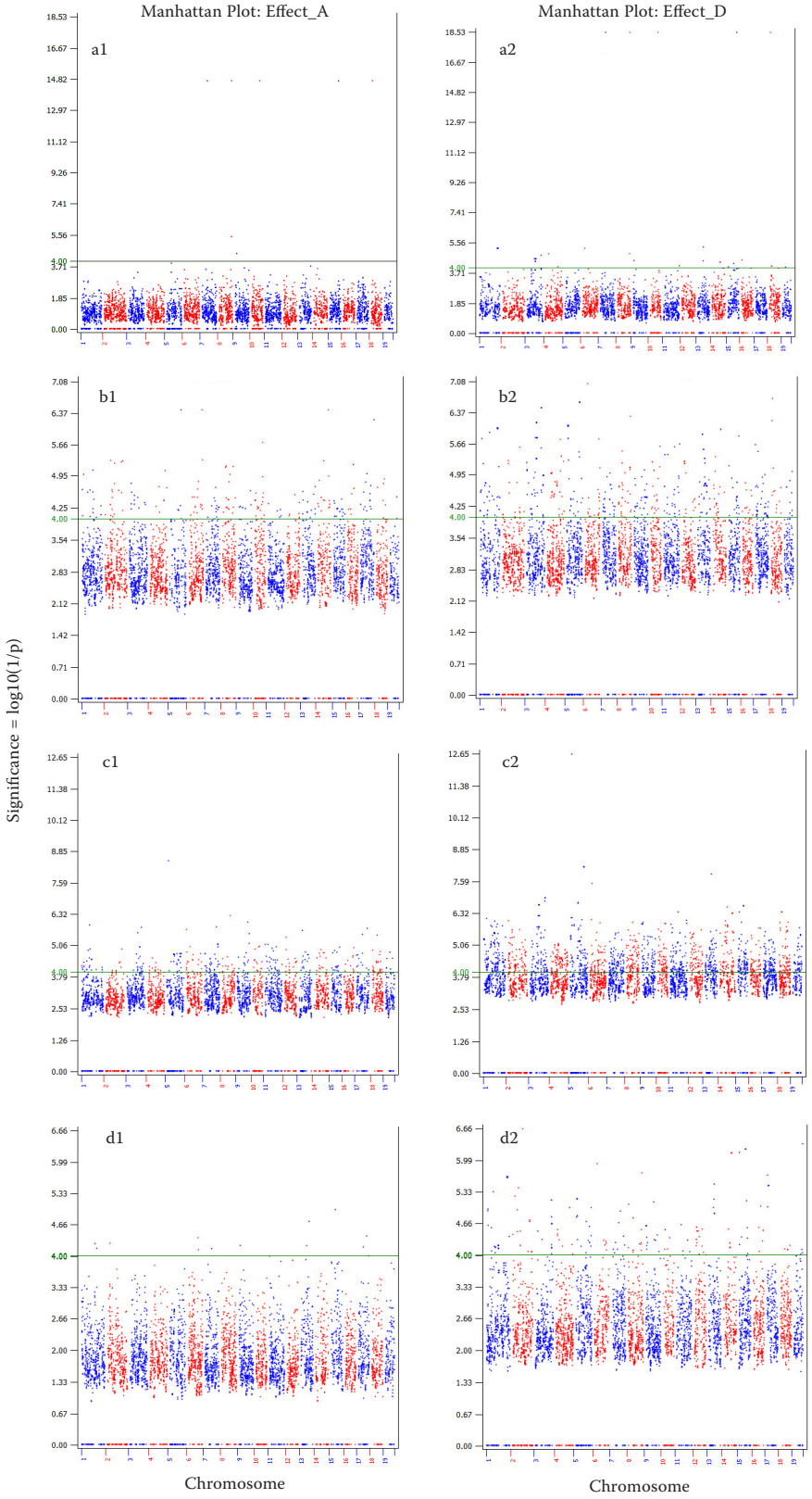

Figure 3. Manhattan plots of additive and dominance effects

a – end weight, b – growth slope, c – body mass index, d – body length; numbers 1 and 2 are for additive and dominance effects, respectively; effects estimated by a model considering both additive and dominance (MAD model); *x*-axis shows chromosome number and *y*-axis shows –log (*P*-value)

https://doi.org/10.17221/83/2017-CJAS

weight, in GVCBLUP-MAD 12, 100, and 7 SNPs had significant additive, dominance, and both effects, respectively ($P < 10^{-4}$). The number of SNPs with a significant additive effect on all traits using GVCBLUP-MAD was slightly lower than the number of SNPs detected by GVCBLUP-MA. However, GVCBLUP-MAD and MD performed differently in detecting SNPs with dominance effects on traits. The number of significant SNPs with a dominance effect resulting in GVCBLUP-MD was lower if compared to GVCBLUP-MAD. Analysis and comparison of residual variances in the three models (MAD, MA and MD) for all traits showed that MAD estimates smaller residual variance for end weight (3.68 vs 4.6) and body length (0.018 vs 0.026) compared to MD ($P < 0.01$), therefore the discovery power in MAD is higher, while the variances explained in MAD and MA are close and so are the discovery powers. Some biological evidence from the National Center for Biotechnology Information databases supports the GWAS findings of the present authors associated with the traits. Tests of association may be based on individual SNPs, but not on the sets of neighbouring SNPs. For example, rs8255002, known as one of top ten SNP markers (additive) significantly associated with end weight, is a C/T single-nucleotide variation on mouse chromosome 8: 121999608, *Banp* gene, which is associated with mouse muscles.

The number of SNPs, that simultaneously showed both additive and dominance effect on traits, was detected to be 498 of 6340 significant SNPs. Also, the results revealed that the numbers of SNPs controlling growth slope and BMI across all models were greater than those controlling end weight and body length ($P < 10^{-4}$). These results, and considering the average of all significant SNPs effects related to each trait, show that these two traits tend to be controlled by greater numbers of SNPs with a smaller effect if compared to end weight and body length.

Based on additive and dominance SNP effects estimated by all models for all traits, ten SNPs with the largest effect were characterised. All significant SNPs and top largest additive and dominance SNPs by chromosome, traits and models are presented in Supplementary Tables S1 to S8 – for supplementary material see the electronic version. The first five SNPs with the largest additive and dominance effects detected by MAD and MA models for end weight were the same and are located on chromosomes 7, 8, 10, 15, 18, respectively. In addition, according to the number and type of significant SNPs, the performance of models in detecting the same SNPs affecting the traits was investigated. The numbers of common significant additive SNPs in MAD and MA models for end weight, growth slope, BMI, and body length were 5, 25, 30, and 45, respectively. Similarity common significant dominance SNPs between MAD and MD models for end weight, growth slope, BMI, and body length were reported to be 25, 30, 42, and 56, respectively.

Statistical significance of the association of SNPs was determined based on Bonferroni-corrected *P*-values. This method treats individual tests as independent and thus is very conservative. Aiming for an overall false positive rate of 0.05 and considering 12 500 independent tests, the pointwise *P*-value should be between 0.0001 and 0.0004. In this article, individual SNPs with *P*-value of 0.0001 or less were considered statistically significant. In order to identify common significant SNPs among traits, a pairwise comparison between traits was performed (Table 4). Several SNPs on chromosome 17 had both large additive and dominance effects for end weight. Three SNPs on chromosome 17 had large dominance and ad-

Table 3. Significant single nucleotide polymorphisms (SNPs) in different traits and models ($P < 10^{-4}$)

Models	MAD			MA			MD		
Traits	A	D	A&D	A	D	A&D	A	D	A&D
End weight	12	100	7	13	–	–	–	63	–
Growth slope	424	1120	126	343	–	–	–	586	–
Body mass index	677	3490	357	679	–	–	–	1240	–
Body length	36	481	8	33	–	–	–	208	–

MAD = SNP marker model considering additive and non-additive (dominance) effects, MA = SNP marker model considering additive effects, MD = SNP marker model considering just dominance effect; A, D and A&D are the numbers of SNPs that had significant additive, dominance, and both effects, respectively

https://doi.org/10.17221/83/2017-CJAS

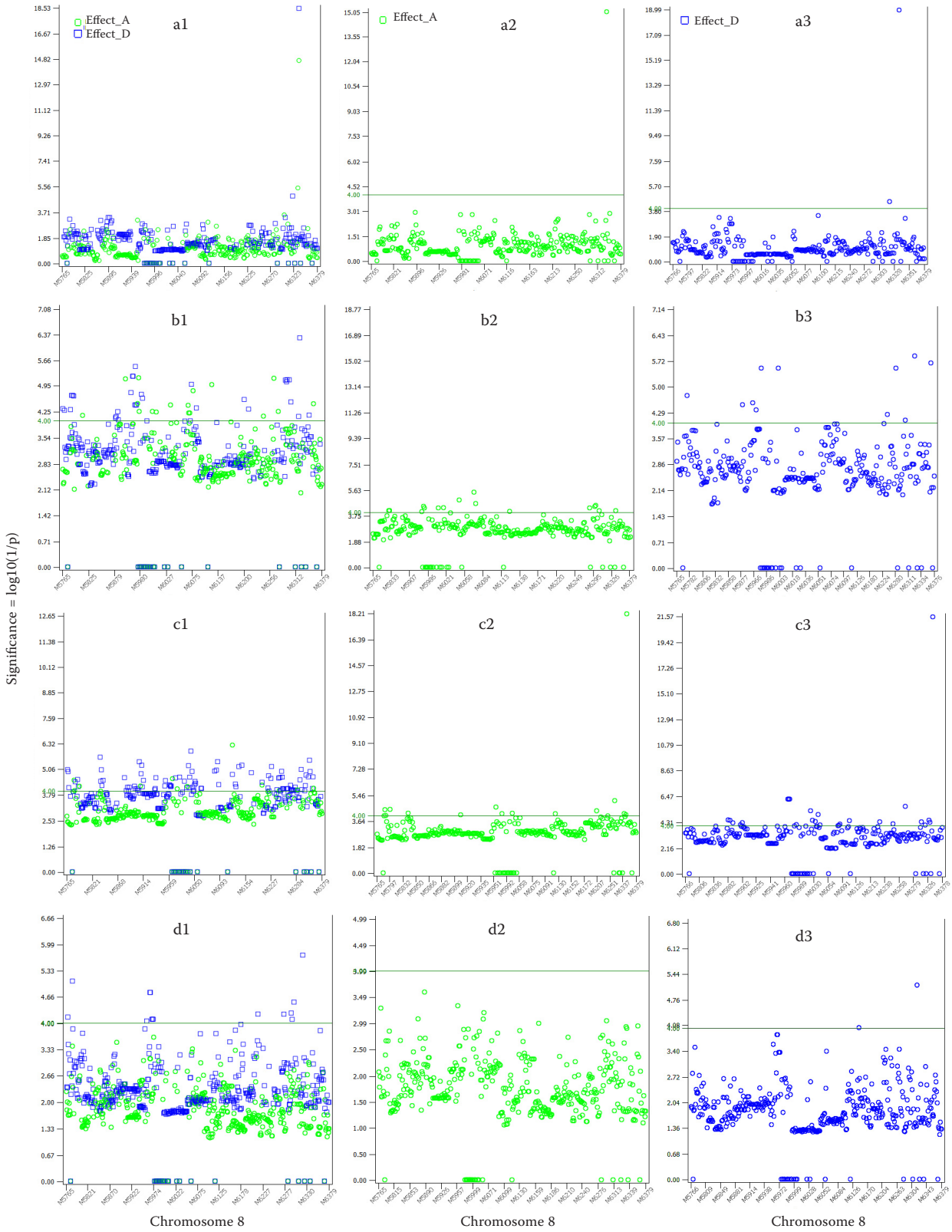

Figure 4. Additive and dominance effects of chromosome 8

a – end weight, b – growth slope, c – body mass index, d – body length; numbers 1, 2, and 3 show additive, dominance, and additive and dominance model, respectively, by which SNP effects were detected; *x*-axis shows SNPs number over the chromosome 8 and *y*-axis shows –log (*P*-value)

https://doi.org/10.17221/83/2017-CJAS

Table 4. A pairwise comparison between traits to identify common significant single nucleotide polymorphisms (SNPs) ($P < 0.0001$)

Traits	End weight	Growth slope	Body mass index	Body length
End weight		2	4	0
Growth slope	81		46	4
Body mass index	83	679		8
Body length	73	298	370	

above diagonal: the number of additive common SNPs; below diagonal: the number of dominance common SNPs

ditive effects for end weight and growth slope. Chromosomes 7, 8, and 10 each had one SNP with a large dominance and additive effect for end weight, and similarity chromosome 5 had one SNP with a large additive and dominance effect for end weight, BMI, and body length. The most common chromosome 8 had the largest additive and dominance effect on all traits. Figure 4 illustrates chromosome 8 SNPs effects detected by the three models. The growth slope and BMI had 46 common additive SNPs that show this trait tends to be correlated more additively than the other traits. Also, it could be concluded that BMI and body length are correlated with other traits through non-additive genetic effects. Generally, SNPs with large additive effects were easily detected by both MAD and MA model so that they were common in both models. But SNPs with small effect were determined differently by the models.

The genomic best linear unbiased predictions including additive, dominance, and both genetic effects termed $GBLUP_A$, $GBLUP_D$, and $GBLUP_G$, respectively, are in Table 5. Reliabilities of estimated additive GBLUPs in different traits ranged from 44.8 to 66.6% for the MAD model, and 46.1 to 69% for the MA model, respectively. These reliabilities were also estimated for dominance GBLUPs and ranged from 6 to 26.4% for the MAD model and 22.5 to 50.5% in the MA model. The MA model led to the higher reliability for estimated additive GBLUPs compared to MAD. The same pattern was observed in reliability for estimated dominance genetic effect, that is, reliability estimated by MD was higher than that estimated by the MAD model. Results showed that differences in reliability of $GBLUP_A$ estimated using MAD and MA for all traits were not significant, except for growth slope ($P < 0.01$). Estimations of $GBLP_D$s despite low reliability for all traits were different across all models. A much larger and significant reliability was only observed in growth slope, a situation in which there is a relatively high dominance to additive effect ratio. Given that a high proportion of genetic variance is additive and to comply with the $GBLUP_A$ estimations, $GBLUP_G$s were not different across models for all traits except for growth slope. These and the highest estimation for $GBLUP_D$ could be due to that growth slope is more affected by dominance compared to other traits. In our study, considering estimated non-additive genetic variances, the gain in the reliability of genomic predictions by including non-additive genetic effects in the prediction model was almost nothing. These results were in contrast to Su (2012) and Zeng (2013) that have shown, even when purely additive effects were evaluated, the inclusion of dominance in the genomic evaluations did not decrease the accuracy of prediction. Others have recently reported that the prediction of dominance deviation from SNP information is not as accurate as that reported for breeding values (Nishio and Satoh 2014). However, the use of larger training populations (Wittenburg et al. 2015) or the adoption of training populations where loci with higher

Table 5. Reliability of GBLUPs in different traits and models

Models	MAD			MA			MD		
Traits	$GBLUP_A$	$GBLUP_D$	$GBLUP_G$	$GBLUP_A$	$GBLUP_D$	$GBLUP_G$	$GBLUP_A$	$GBLUP_D$	$GBLUP_G$
End weight	0.666	0.196	0.655	0.69^{ns}	–	0.69^{ns}	–	0.436	0.436**
Growth slope	0.604	0.264	0.599	0.654**	–	0.654**	–	0.505	0.505**
Body mass index	0.448	0.06	0.425	0.461^{ns}	–	0.461^{ns}	–	0.225	0.225**
Body length	0.567	0.114	0.547	0.593^{ns}	–	0.593*	–	0.312	0.312**

MAD = SNP marker model considering additive and non-additive (dominance) effects, MA = SNP marker model considering additive effects, MD = SNP marker model considering just dominance effect, GBLUPs = Genomic best linear unbiased prediction; A, D and G show GBLUPs estimated by considering additive, dominance, and both effects

*$P < 0.05$, **$P < 0.01$, ns = not significant

https://doi.org/10.17221/83/2017-CJAS

minor allele frequency occur (and therefore more heterozygotes are available for dominance estimation) may improve predictions.

In order to evaluate the performance of models in ranking animals using GBLUPs, correlations (Spearman) between synonymous GBLUPs estimated by different models have been calculated. Correlations between $GBLUP_A$ estimated by MAD and MA models for all traits were high and ranged from 0.982 to 0.997. Similarity correlations between $GBLUP_D$ estimated by MAD and MD models were calculated. These correlations were lower than those for $GBLUP_A$s and ranged from 0.687 to 0.898. MAD and MA models had a similar performance in ranking animals according to estimated $GBLUP_G$s and comparing MAD vs MD model and MA vs MD model showed a moderate similarity in ranking animals.

However, including dominance in models could still be beneficial, for example, in predicting disease risk in humans (Wray et al. 2007) or for establishing mating strategies in plant or animal breeding aimed at maximising the phenotypic performance of the (crossbred) offspring. Also, the incorporation of dominance effects is critical for the introduction of breeding approaches that aim to create crosses with complementary alleles in mate-pair allocation (Munoz et al. 2014). In addition, using dominance in genomic evaluations is expected to result in greater cumulative response to selection of purebred animals for crossbred performance than additive models, especially in the presence of overdominance and when retraining is not performed at each generation (Zeng et al. 2013).

Compared to the pedigree-based relationship matrix, the genomic relationship matrix can capture both Mendelian segregation and the genetic links through unknown common ancestors, which are not available in the known pedigree. Furthermore, the genomic relationship matrices are applicable for different populations with or without pedigree information, which is particularly advantageous in studies on wild populations or human populations (Visscher et al. 2010).

The lack of improvement by including dominance effect using SNP markers in some traits and slightly better performance in other traits indicates the difficulty in distinguishing additive and non-additive genetic effects. Also, it could indicate that traits with a large variance of additive effects were influenced little by the inclusion of dominance, but greatly for the traits with a lot of dominance effects.

CONCLUSION

Using SNPs dense markers and phenotype solely without pedigree information provides a new approach to detect additive and dominance genetic variation and predict genetic merit. This study evaluated a GBLUP framework, and orthogonal for using marker-based information in genomic predictions involving both additive and non-additive effects. Also, the current and previous studies have shown that non-additive genetic variance is remarkable in complex traits. The advantage of using SNPs marker is that fitting a genome-enabled prediction model on just SNP information is beneficial especially for species in which there is a lack or unreliable pedigree information. Using SNPs and including non-additive in models despite decreasing the reliability of GBLUPs estimation could still be beneficial for establishing mating strategies in plant or animal breeding aimed at maximising the phenotypic performance of the crossbred offspring. Thus using models unbiased estimation of total genetic value that includes additive and non-additive effects can be an effective tool for predicting an individual and future offspring's total genetic potential and phenotype. This may lead to more genetic progress in selection and mating systems. The beneficiality of using SNPs and including non-additive in models depends on the architecture of traits and different heritability of traits. In human context, considering results of this study, following significant SNPs affecting diabetes-related traits by post-genome wide studies could reveal unknown aspects associated with genetical control of the disease.

Acknowledgement. We gratefully acknowledge the Wellcome Trust Centre for Human Genetics, Oxford, for making the heterogeneous stock data available at http://mtweb.cs.ucl.ac.uk/mus/www/mouse/HS/index.shtml.

REFERENCES

Alvarez-Castro J.M., Carlborg O. (2007): A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. Genetics, 176, 1151–1167.

Armando C., Albert T., Peter D.K. (2015): The nature of genetic variation for complex traits revealed by GWAS and Regional Heritability Mapping analyses. Genetics, 201, 1601–1613.

Carlborg R., Hocking P.M., Burt D.W., Haley C.S. (2004): Simultaneous mapping of epistatic QTL in chickens re-

https://doi.org/10.17221/83/2017-CJAS

veals clusters of QTL pairs with similar genetic effects on growth. Genetical Research, 83, 197–209.

de los Campos G., Hickey J.M., Pong-Wong R., Daetwyler H.D., Calus M.P. (2013): Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics, 193, 327–345.

Duangjinda M., Bertrand J.K., Misztal I., Druet T. (2001): Estimation of additive and nonadditive genetic variances in Hereford, Gelbvieh, and Charolais by Method R. Journal of Animal Science, 79, 2997–3001.

Grosse-Brinkhaus C., Jonas E., Buschbell H., Phatsara C., Tesfaye D., Jungst H., Looft C., Schellander K., Tholen E. (2010): Epistatic QTL pairs associated with meat quality and carcass composition traits in a porcine Duroc × Pietrain population. Genetics Selection Evolution, 42, 1–13.

Hill W.G., Goddard M.E., Visscher P.M. (2008): Data and theory point to the mainly additive genetic variance for complex traits. PLoS Genetics, 4, 1000008.

Jiang Y., Reif J.C. (2015): Modeling epistasis in genomic selection. Genetics, 201, 759–768.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., Lee D.H. (2002): BLUPF90 and related programs (BGF90). In: Proc. 7th Annu. World Congress on Genetics Applied to Livestock Production, Montpellier, France.

Mott R., Christopher J., Talbot M.G., Turri A.C., Collins J.F. (2000): A method for fine mapping quantitative trait loci in outbred animal stocks. Proceedings of the National Academy of Sciences of the United States of America, 97, 12649–12654.

Munoz P.R, Resende M.F.J., Gezan S.A., Resende M.D., de los Campos G., Kirst M., Huber D., Peter G.F. (2014): Unraveling additive from non-additive effects using genomic relationship matrices. Genetics, 198, 1759–1768.

Nishio M., Satoh M. (2014): Including dominance effects in the genomic BLUP method for genomic evaluation. PLoS ONE, 9, e85792.

Palucci V., Schaeffer L.R., Miglior F., Osborne V. (2007): Non-additive genetic effects for fertility traits in Canadian Holstein cattle. Genetics Selection Evolution, 39, 181–193.

Su G., Christensen O.F., Ostersen T., Henryon M., Lund M.S. (2012): Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. PLoS ONE, 7, e45293.

Tempelman R.J, Burnside E.B. (1990): Additive and non-additive genetic variation for conformation traits in Canadian Holsteins. Journal of Dairy Science, 73, 2214–2220.

Valdar W., Solberg L.C., Gauguier D., Burnett S., Klenerman P., Cookson W., Taylor M., Rawlins N.P., Mott R., Flint J. (2006): Genome-wide genetic association of complex traits in heterogeneous stock mice. Nature Genetics, 38, 879–887.

Van Tassell C.P., Misztal I., Varona L. (2000): Method R estimates of additive genetic, dominance genetic, and a permanent environmental fraction of variance for yield and health traits of Holsteins. Journal of Dairy Science, 83, 1873–1877.

Vinkhuyzen A.A.E., Wray N.R., Yang J., Goddard M.E., Visscher P.M. (2013): Estimation and partitioning of heritability in human populations using whole genome analysis methods. Annual Review of Genetics, 47, 75–95.

Visscher P.M., Yang J.A., Goddard M.E. (2010): A commentary on "common SNPs explain a large proportion of the heritability for human height" by Yang et al. (2010). Twin Research and Human Genetics, 13, 517–524.

Vitezica Z.G., Legarra A., Toro M.A., Varona L. (2017): Orthogonal estimates of variances for additive, dominance and epistatic effects in populations. Genetics, 206, 1297–1307.

Wang C., Dzianis P., Shengwen W., Sujata P., Hakizumwami B.R., Yang D. (2014): GVCBLUP: a computer package for genomic prediction and variance component estimation of additive and dominance effects. BMC Bioinformatics, 15, 1–13.

Wang S., Dvorkin D., Da Y. (2012): SNPEVG: A graphical tool for GWAS graphing with mouse clicks. BMC Bioinformatics, 13, 1–6.

Wellmann R., Bennewitz J. (2012): Bayesian models with dominance effects for genomic evaluation of quantitative traits. Genetics Research, 94, 21–37.

Wittenburg D., Melzer N., Reinsch N. (2015): Genomic additive and dominance variance of milk performance traits. Journal of Animal Breeding and Genetics, 132, 3–8.

Wood A., Esko T., Yang J. et al. (2014): Defining the role of common variation in the genomic and biological architecture of adult human height. Nature Genetics, 46, 1173–1186.

Wray N.R., Goddard M.E., Visscher P.M. (2007): Prediction of individual genetic risk to disease from genome-wide association studies. Genome Research, 17, 1520–1528.

Yang D., Wang C., Wang S., Hu G. (2014): Mixed model methods for genomic prediction and variance component estimation of additive and dominance effects using SNP markers. PLoS ONE, 9, e87666.

Zeng J., Toosi A., Fernando R.L., Dekkers J.C., Garrick D.J. (2013): Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. Genetics Selection Evolution, 45, 1–17.

Zhang W., Dai X., Wang Q., Xu S., Zhao P.X. (2016): PEPIS: A pipeline for estimating epistatic effects in quantitative trait locus mapping and genome-wide association studies. PLOS Computational Biology, 12, e1004925.

Received: 2017–07–19
Accepted after corrections: 2018–09–11