Exploring non-additive variance using genome-wide dense SNP in four French and Nordic dairy cattle breeds

A. Marete[1,2], L. Janss[,2], B. Guldbrandtsen[2], Hoze C[1,3], G. Sahana[2], D. Boichard[1], M. S. Lund[2]

[1]GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France
[2]Aarhus University, Center for Quantitative Genetics and Genomics, Denmark
[3]Allice, 75595 Paris, France
andrew.marete@mbg.au.dk (Corresponding Author)

Summary
Non-additive genetic variation is usually ignored in studies of the genetic architecture and genomic prediction of complex traits. However, non-additive genetic effects may have an important contribution to total genetic variation. This study presented a Bayesian mixed model including additive and non-additive genetic effects, with corresponding genetic relationship matrices obtained from genome-wide dense SNPs. The amounts of additive genetic and dominance variance were estimated from all SNPs whereas epistatic variance was estimated for a set of pre-selected markers with a major effect. Variances were estimated for milk, fat, and protein yields from 54,862 cows in four dairy populations. In general, variation due to non-additive effects was present in all the four breeds with Montbeliarde and Holstein breeds having more dominance and epistatic variation. Dominance variance was quite consistent across breeds and traits and represented about 20% of the additive genetic variance. The epistatic variance estimates were more variable, from nearly zero to 19% of the additive genetic variance, with an average of 7%.

Keywords: non-additive, epistasis, SNP-by-SNP interaction, bovine, genomics, WGS

Introduction

Complex traits, such as milk production, are regulated by a complex interplay between multiple genes, each with a small effect, and environmental factors (Mackay, 2014). These genes, theoretically, can interact with each other. This interaction can result in partition of genetic variance into additive variance, and non-additive variance. Non-additive variance is composed of interaction between genes at the same locus (dominance deviation) or interactions between genes at different loci (epistatic variance) (Falconer, 1975). Over the years, however, most genetics studies have relied heavily on the additive component, because it is believed to be the largest component, easiest to estimate, and easiest to use in selection. Consequently, there is limited amount of knowledge on how much genetic variation is explained by the non-additive variation in dairy cattle; in production traits, it is still largely unknown whether accounting for both additive effects and non-additive effects improves the accuracy of prediction.

Main limiting factors for answering this question was lack of genotyped individuals with own performance records, and lack of methods allowing genomic models to efficiently account for non-additive effects. If contribution of non-additive effects to the phenotypic variance is substantial, including non-additive effects in genetic evaluation models could improve breeding values estimation. It could also lead to a more accurate prediction of individual's future performance. In this study, we explore the variations due to additive effects, dominance deviation and 2x2 epistasis interactions in four dairy cattle populations.

Materials and method

Study population

Cows were genotyped with either Illumina BovineSNP50 beadchip, or the customized low density EuroGenomics SNP chip (Boichard et al., 2018, this conference). The EuroGenomics SNP chip is composed of two parts: (1) ~10,000 generic (and supposedly neutral) SNP mostly from the bovine LD chip (Boichard et al, 2012); and (2) a custom part selected from whole genome sequence based on five functional arguments: (i) variants described in literature (eg. K232A mutation in DGAT1), (ii) potential regulatory variants located in the promoters of some genes, (iii) non-synonymous variants with strongly deleterious effect on the function of the encoded protein as predicted by Variant Effect Predictor (VEP) (McLaren et al., 2016), (iv) breakpoints of structural variants affecting genes as described in (Boussaha et al., 2015), and (v) variants corresponding to GWAS peaks. All SNPs with MAF lower than 0.5%, with a call rate lower than 95%, or deviating from Hardy-Weinberg equilibrium were deleted. Four versions of this Eurogenomics chip were used, with partial overlap between custom parts. To get a complete marker information across individuals, imputation was carried out within breed using FImpute (Sargolzaei et al, 2014), considering all genotyped animals (males and females, with or without performances). After imputation, an average of 49,835 SNPs distributed over 29 *Bos taurus* autosomes (BTA) remained.

Phenotypic records were yield deviations for milk, fat and protein yields. They were obtained from French and Nordic evaluation systems. The number of genotyped animals with phenotypes was: Montbeliarde: 19,788, Normande: 11,978, Danish-Jersey: 10,538, Holstein: 12,558, summing up to 54,862 cows.

Statistical Models

Bayz software (www.bayz.biz) was used to estimate variance components and SNP effects

i. Dominance (ad model)

The model used for estimating dominance was the following:

$\mathbf{y = 1}\mu + \mathbf{Xa + Zd + e}$

where: **y** is a vector of yield deviations, already adjusted for non-genetic effects, μ is a mean, **X** and **Z** are the (p,n) matrices of genotypes relative to additive and dominance effects, p the number of individuals and n the number of SNP; **a, d**, and **e** vectors of additive, dominance and residual effects, which are assumed to be normally distributed with 0 mean and variance $\mathbf{I}\ \sigma_a^2$, $\mathbf{I}\ \sigma_d^2$, and $\mathbf{I}\ \sigma_e^2$. General terms of **X** and **Z** are x_{ik} = -1,0, and 1, and z_{ik} = 0, 1, and 0 for AA, AB and BB genotypes, respectively. Narrow sense heritability was estimated as and dominance deviation as where .

ii. SNP-by-SNP interaction

a) Selecting SNPs to form interaction matrix

To limit dimension of analysis, 2x2 epistasis was studied only for SNPs with a major effect on the phenotype of at least one breed. These SNPs were identified by GWAS using GCTA software (Yang et al, 2011), accounting for polygenic effects of each individual. Corresponding genomic matrices were built with 43,800 SNPs. After Bonferroni correction, SNPs were selected with a genome-wide significance threshold of 10^{-6}. The most significant SNP was selected in a 2Mb window.

b) The (aa) model:

Across the four breeds, 237 unique GWAS peaks selected formed the interaction matrix (epistatic SNP effects). The model fitted additive and epistatic effects as follows:

$\mathbf{y} = \mu + \mathbf{Xa + Wc + e}$

with **y**, μ, **X, a** and **e** defined as above. **c** is the vector of epistatic effects and **W** is a matrix of

genotypes with respect to epistatic effects. For individual i and combination of variants (j,k), the general term of **W** is the following

wij	Locus k		
Locus j	AA	AB	BB
AA	1	0	-1
AB	0	0	0
BB	-1	0	1

Distributions of both **a** and **c** are mixtures of two distributions, allowing for a large or a small effect.

where the upper term is distribution of SNPs with smallest effects, and the lower term is distribution of SNPs with largest effects; For each iterate, the model estimates were a result of a Monte-Carlo chain with Metropolis-Hastings sampler. For both models (ad and aa), quality of fit was assessed by Monte-Carlo coefficient of variation (MCCV)

The model was assumed to be fit if MCCV (%) < 1. For **aa** model, in addition, the acceptance rate of the Metropolis-Hastings sampling was considered acceptable if it fell within [0.2:0.8] interval.

Results

Lowest MCCV was observed for additive variance estimates in both SNP-by-SNP (**aa**) and additive-dominance (**ad**) models (≤0.1%). In aa-model, non-additive variance estimates had varying MCMC (4.8% – 5.4%), depending on trait and breed size. In **ad**-model, non-additive effects MCCV were <1% for all traits and breeds.

Table 1 shows posterior means of variance components for both **ad**- and **aa**-models. For most breeds and traits, dominance deviation reached 18% – 25% of the additive variance. Jersey had lowest ratio of dominance deviation to additive variance (10.4%).

The **ad**-model had a higher additive variance component compared to **aa**-model. Results were variable according to breed. Epistasis over additive variance varied from 15% – 19% in Montbeliarde breed to 0.7% – 1.4% in Jersey breed. Normande and Holstein had intermediate results with ratios ranging from 4.8% – 8.8%.

Table 1. Posterior means of residual variances (), additive genetic variance () and dominance deviation ()

Trait	Parameter	Mean variance (standard error) for breed			
AD Model					
		Montbeliarde	Normande	Jersey	Holstein
Fat		771 (9)	744 (14)	1179 (21)	1246 (23)
		227 (8)	328 (14)	463 (20)	560 (18)
		43 (6)	77 (10)	96 (15)	90 (16)
Milk (x 1000)		455 (6)	337 (6)	368 (7)	818 (15)
		140 (4)	174 (7)	275 (8)	471 (15)
		26 (3)	36 (5)	29 (5)	78 (11)
Protein		529 (6)	410 (8)	552 (10)	826 (15)
		129 (5)	183 (8)	239 (10)	351 (12)
		33 (4)	40 (6)	46 (7)	76 (10)
AA Model					
Fat		793 (10)	787 (14)	1248 (20)	1310 (21)
		205 (9)	328 (15)	486 (19)	552 (23)

		34 (6)	29 (9)	5 (5)	33 (14)
Milk (x1000)		469 (6)	358 (6)	390 (7)	871 (14)
		125 (6)	178 (7)	280 (8)	470(16)
		18.4 (3)	9.1 (4)	2.1 (2)	22.6 (9)
Protein		546 (7)	434 (7)	584 (10)	877 (14)
		115 (6)	184 (9)	250 (10)	357 (15)
		22 (4)	12 (4)	4 (3)	18 (9)

Heritability coefficients are presented in table 2. Narrow sense heritability matches with literature data, from 0.19 – 0.41 according to breeds and traits. As expected, values were highest for milk and lowest for protein. Across breeds, highest values were obtained in Jersey and lowest in Montbéliarde. Results were similar in **aa**- and **ad**-models.

Dominance heritability varied from 4.1% – 6.7%, with lowest values in Montbéliarde and highest in Normande breeds. Results varied more according to breeds than to traits.

Heritable portion explained by SNP-by-SNP interactions varied from 0.3% – 3.3%. Results varied mainly according to breeds, with highest values observed in Montbeliarde and lowest in Jersey. Within breed, difference between traits were limited.

Table 2. Heritability estimates with standard errors for three production traits (x 100) in the four breeds

Trait	Parameter	Breed			
		Montbeliarde	Normande	Jersey	Holstein
Model 1: Additive-dominance model					
Fat		21.8 (0.7)	8.6 (1.0)	26.6 (1.0)	29.5 (0.8)
		4.1 (0.6)	6.7 (0.9)	5.5 (0.8)	4.7 (0.9)
Milk		22.5 (0.6)	31.8 (1.0)	41.0 (0.9)	34.5 (1.0)
		4.2 (0.6)	6.6 (0.9)	4.3 (0.7)	5.6 (0.8)
Protein		18.6 (0.7)	28.9 (1.1)	28.5 (0.9)	28.0 (0.8)
		4.7 (0.6)	6.3 (1.0)	5.5 (0.9)	6.1 (0.8)
Model 2: SNP-by-SNP model					
Fat		19.9 (0.8)	28.7 (1.2)	28.0 (0.9)	29.2 (1.0)
		3.3 (0.6)	2.5 (0.7)	0.3 (0.3)	1.8 (0.7)
Milk		20.4 (0.8)	32.7 (1.2)	41.6 (0.9)	34.5 (1.0)
		3.0 (0.6)	1.7 (0.7)	0.3 (0.3)	1.7 (0.7)
Protein		16.9 (0.8)	29.2 (1.2)	29.8 (1.0)	28.5 (1.1)
		3.2 (0.6)	1.8 (0.7)	0.4 (0.4)	1.5 (0.7)

Discussion

Estimated dominance variance (V_d) in proportion to additive genetic variance (V_a) averaged over three traits was ≈20%. Estimates of epistasis variance (V_{ep}) were lower than Vd and more variable, ranging from nearly 0 –19% of additive component, with an average of 7%.

Current dominance results are in range of those for complex traits reported in previous studies (Su et al, 2012). In dairy cattle, the ratio of V_d to V_a was 17% for stature in US Holstein (Misztal, 1997). In beef cattle, ratio of V_d to V_a was larger than 50% for weaning weight in Hereford, Gelbvieh and Charolais beef cattle (Duangjinda et al, 2001; Gengler et al, 1997). Ratios of V_d to V_a ranged from 11% – 31% for other traits such as reproductive and growth traits in Yorkshire pigs (Culbertson et al., 1998). These results indicate dominance variations are important for complex traits, although clearly much lower than additive components.

We estimated V_{ep} using a combination of genome-wide SNP markers with a major effect. This strategy reduces dimensionality of analysis to about 28,000 SNP combinations and limits linkage disequilibrium between SNPs. This strategy is only an approximation, as strong epistatic

effects may exist without additive effect. However, this situation is quite rare and occurs only with so-called "sign" epistasis, whereas most epistatic effects also generate additive effects. Therefore, focusing on markers with significant additive effects should capture largest part of epistatic effects. The use of a stringent threshold may have resulted to the loss of large epistatic interactions. Our estimate was variable, with epistatic heritability coefficients ranging for 0.3% – 3.3% according to breed and trait. This variability illustrates the difficulty to estimate epistasis despite reasonably large genotyped populations. This variability can be explained by lack of informativity of some SNPs, especially for Jersey. The V_{ep} when estimated with a limited number of SNPs appears to be negligible for practical selection. V_{ep} however would be valuable in the prediction of the future performances of individuals.

Acknowledgements
The authors are grateful to the French National Animal Breeding Database for providing phenotypic data, Valogene for access to genotypes and INRA for providing the study environment. A. Marete is the recipient of an Erasmus Mundus Ph.D. grant funded jointly by the European Graduate School in Animal Breeding and Genetics (France), Viking Genetics (Denmark) and the Center for Genomic Selection in Animals and Plants (GenSAP) funded by Innovation Fund Denmark (grant 0603-00519B).

List of References

Boichard, D., Chung, H., Dassonneville, R., David, X., Eggen, A. A., Fritz, S. S., et al. 2012. Design of a bovine low-density snp array optimized for imputation. Plos One, *7*, e34130.

Boussaha, M., Esquerré, D., Barbieri, J., Djari, A., Pinton, A., Letaief, R., et al. 2015. Genome-wide study of structural variants in bovine Holstein, Montbeliarde and Normande dairy breeds. Plos One, 10, e0135931.

Culbertson, M. S., Mabry, J. W., Misztal, I., Gengler, N., Bertrand, J. K., & Varona, L. 1998. Estimation of Dominance Variance in Purebred Yorkshire Swine. JAnim Sci, *76*, 448–451.

Duangjinda, M., Bertrand, J. K., Misztal, L., & Druet, T. 2001. Estimation of additive and nonadditive genetic variances in Hereford, Gelbvieh, and Charolais by method . J Anim Sci, *79*, 2997–3001.

Falconer, D. S. 1975. Introduction to quantitative genetics. Pearson Education India.

Gengler, N., Van Vleck, L. D., MacNeil, M. D., Misztal, I., & Pariacote, F. A. 1997. Influence of Dominance Relationships on the Estimation of Dominance Variance with Sire-Dam Subclass Effects. J Anim Sci, *75*, 2885–2891.

Mackay, T. F. C. 2014. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat Rev Genet, *15*, 22–33.

McLaren, W., B. Pritchard, D. Rios, Y. Chen, P. Flicek, & F. Cunningham. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics, 26, 2069-2070.

Misztal, I. 1997. Estimation of Variance Components with Large-Scale Dominance Models. J. Dairy Sci, *80*, 965–974.

Sargolzaei, M., J. P. Chesnais, & F.S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics, 15, 478.

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*.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. 2011. GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet, *88*, 76–82.