Sequence-based multi-trait genome-wide association study for linear classification traits in Belgian Blue beef cattle

J.L. Gualdron Duarte[1,2*], A.S. Gori[2], G. Costa Monteiro Moreira[1], M. Georges[1], C. Charlier[1] and T. Druet[1]

[1]Unit of Animal Genomics, GIGA-R, University of Liège, 4000-Liège, Belgium; [2]Walloon Breeders Association (awe groupe), 5590-Ciney, Belgium; [*]jlgualdron@awegroupe.be

Abstract

We performed a genome-wide association study in Belgian Blue beef cattle for 11 traits related to muscular development and height. Phenotypes were available for 14,762 genotyped cows and imputation was performed to the sequence level (11,537,240 variants). A total of 37 genome-wide significant QTL were identified with a single-trait analysis, corresponding to 11 unique pleiotropic QTL. We then used and compared two multiple-trait approaches to combine information from different traits to fine-map the QTL and increase the power. Although these approaches allowed identifying additional QTL, other were lost. Among the identified QTL, four corresponded to previously identified variants having negative effect when homozygous. In addition, three new missense variants in genes related to muscular development and height were identified.

Introduction

Genome-wide association studies (**GWAS**) are performed to identify variants associated with traits of interest Generally, they are performed with single-trait (**ST**) approaches although it is not uncommon to identify variants associated with several traits. Multiple-trait (**MT**) approaches could be advantageous in such cases, by improving the fine-mapping resolution or the mapping power. These MT approaches have a high computational cost and they can thus be applied on a limited number of traits. Bolormaa *et al.* (2014) proposed an approximate MT approach that requires only to perform ST approaches for each trait.

With the implementation of genomic selection in Belgian Blue beef (BBB) cattle, the number of genotyped individuals is continuously increasing, offering the possibility to perform GWAS at increasing power. A series of linear classification traits, related to conformation and muscular development, are routinely recorded on adult females. Such traits are typically correlated and affected by shared variants. Our objective was to conduct a GWAS in BBB cattle for 11 linear classification traits and to assess the benefit of MT approaches (Stephens, 2013), including the indirect approach proposed by Bolormaa *et al.* (2014).

Materials & Methods

Genotype Imputation. 14,762 cows with phenotypes were genotyped with custom low-density (LD) arrays with 8,576 to 32,318 SNPs or with the Bovine50K genotyping array (MD). To impute genotypes to the sequence level, we relied on a set of 890 bulls genotyped with the BovineHD array and on a sample of 242 whole-genome sequenced bulls. The imputation was achieved in three steps with ShapeIT4.2 (Delaneau *et al.*, 2019) and Minimac4 (Howie *et al.*, 2012), from LD (7,525 SNPs) to MD (28,893 SNPs), from MD to HD (611,322 SNPs) and from HD to whole-genome sequence level (11,537,240 selected SNPs and indels).

Single-trait GWAS. Association was tested with a Linear Mixed Model (LMM) approach implemented in GEMMA (Zhou and Stephens, 2012). The genomic relationship matrix

obtained with the MD genotypes was used to capture the polygenic effects and the population structure. Each cow was phenotyped for 11 traits including body length, pelvis length, height, chest width, pelvis width, rib shape, rump, top muscling, shoulder muscling and buttock muscling (side and rear view). Phenotypes were corrected for fixed effects from the genetic evaluation. The number of independent SNPs was estimated to be equal to 500,000 whereas we determined that 7 independent traits were analysed. Accordingly, we set the significance threshold at 1.43e-8 after application of a Bonferroni correction for 3,500,000 tests.

Multiple-trait GWAS. We first applied the MT approach proposed by Stephens (2013) and implemented in GEMMA. In addition, we used the indirect approach proposed by Bolormaa *et al.* (2014) relying on a score computed from the signed t-values obtained for the SNPs and for each traits (in ST analyses). For one SNP, the score is computed as **tVt'** where **t** is the array of t-values estimated for the tested SNP for each of the N traits and **V** is the N x N correlation matrix between t-values, estimated from all the tested positions. This score is approximately distributed as a chi-squared with N degrees of freedom. For MT analysis, the significant thresholds were modified to 1e-7 and 5e-8 when one or two genome scans were performed (e.g., all traits simultaneously or two groups of six traits).

Figure 1. Heat-map showing association between QTL and traits. The circle sizes indicate strength of association, red borders around circles are used for significant associations, and the colour intensity denotes effect magnitude.

Results

The ST GWAS analysis identified a total of 37 QTL exceeding the significance threshold. However, several of these QTL co-localized in the same interval, suggesting pleiotropic QTL. All the identified QTL are summarized in Figure 1, where the 37 QTL could be condensed in a set of 11 unique QTL. As illustrated for a QTL on chromosome 19 (Figure 2), high correlations between SNP significance levels for several traits further support the presence of a pleiotropic QTL.

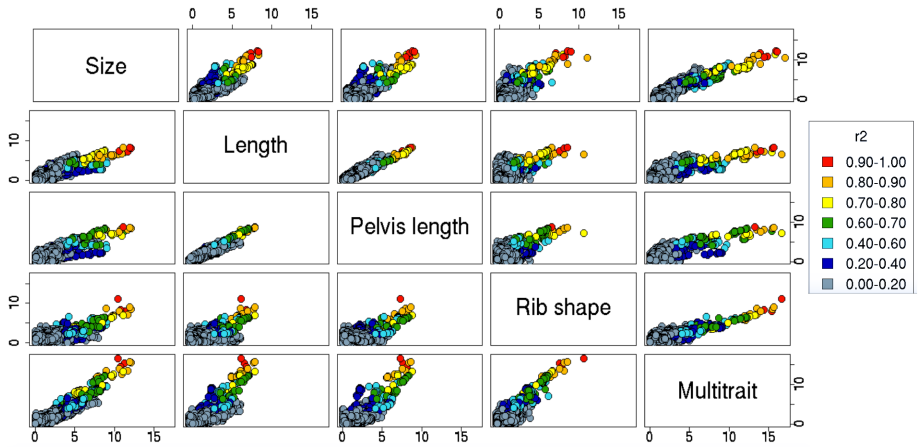

Figure 2. Correlation plot between ST and MT p-values (on -log10 scale) for a QTL on chromosome 19. Variants ± 2.5 Mb from the QTL are included. The dot colours indicate the linkage-disequilibrium (r^2) between each variant and the lead SNP (on each row, the lead SNP corresponds to the analysis indicated on the diagonal).

The first MT approach (Stephens, 2013) could not be applied on all traits simultaneously. We used the results from the QTL sharing with the ST approach to define two groups of 6 traits to perform MT analysis. By combining information from several traits, the MT approach helps to identify the causative variants. For instance, three previously identified causative variants are clearly the lead SNP with the MT approach whereas this was not the case for all ST analyses (Figure 3, Figure 2). The association signals obtained with the second MT mapping approach were highly correlated with those from the first approach (correlations ranging from 0.90 and 0.99) and pointed towards similar sets of causative variants, indicating that the indirect approach was a good approximation (although with slightly lower p-values). Although both approaches identified respectively two and three new QTL, two and one QTL were no longer significant (on BTA16 and 23). Inclusion of traits not associated with a QTL in MT approaches might thus reduce the power of association. Therefore, the optimal MT approach should not necessarily combine all traits simultaneously. These observations were confirmed by a MT analysis including 11 traits: although the mapping resolution was improved for the most pleotropic effects, the power decreased for some QTL associated with fewer traits (QTLs on BTA16 and 23 were lost). Nevertheless, five new QTL were identified thanks to increased power for pleiotropic QTL (and correction for fewer tests).
Four of the QTL are associated to variants altering the protein sequence previously identified, whereas three additional missense variants in genes related to height or muscular development were also identified.

Discussion

Several QTL detected in this study were previously identified in other studies. For instance, we found QTL on BTA5 (CCND2), BTA6 (LCORL and NCAPG), BTA23 and BTA26 reported for stature in Bouwman *et al.* (2018). In addition to these QTL, four QTL associated with deleterious variants presenting heterozygous advantage (f.i., Gualdron *et al.,* 2020), and other QTL might be more specific to the BBB cattle breed. The association with these recessive deleterious variants are still highly significant although their frequency has been decreasing in the population as a result of selection against carriers.
MT approaches allow obtaining a more precise association signal, in particular when most of the combined traits are associated with the variant. Although fewer tests were performed with such approaches, few new QTL were identified and some ST QTL were even lost, possibly

because the MT approach included too many un-associated traits. Finally, the indirect MT approach from Bolormaa *et al.* (2014) was an excellent approximation of the standard MT approach with a lower computational cost. However, the p-values obtained with this second approach were less conservative and it remains to be determined whether the higher number of significant results we obtained results in increased power or a higher rate of false positives

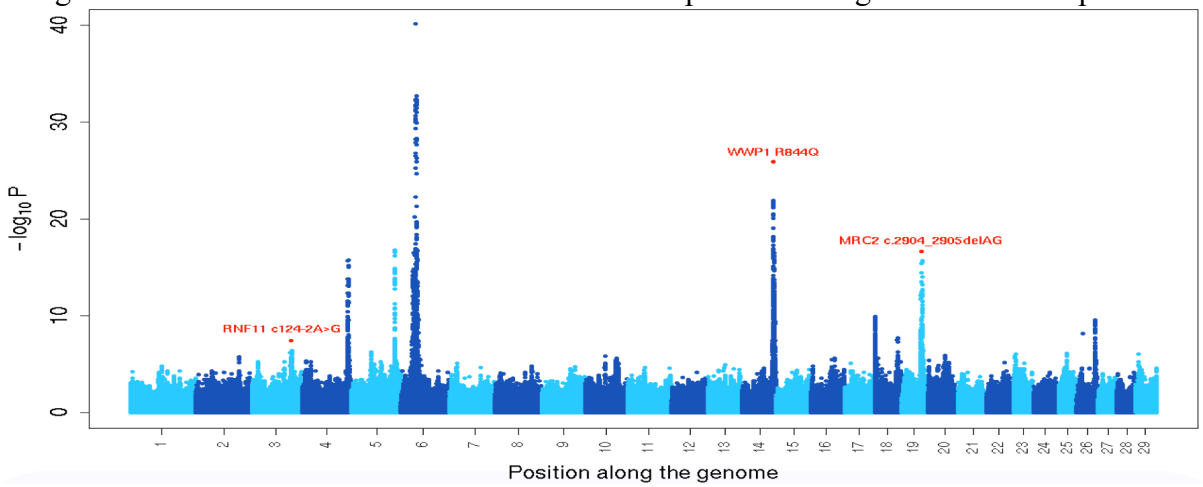

Figure 3. Example of Multiple-trait GWAS. Six traits were included: body length, pelvis length, height, chest width, pelvis width and rib shape. The red points are previously identified causative variants.

References

Bolormaa S., Pryce J.E., Reverter A., Zhang Y., Barendse W. *et al.* (2014) A Multi-Trait, Meta-analysis for Detecting Pleiotropic Polymorphisms for Stature, Fatness and Reproduction in Beef Cattle. PLoS Genet 10(3): e1004198. doi:10.1371/journal.pgen.1004198

Bouwman A.C., Daetwyler, H.D., Chamberlain A.J., Hurtado Ponce C., Sargolzaei M. *et al.* (2018) Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat Genet 50, 362–367. doi.org/10.1038/s41588-018-0056-5

Delaneau O., Zagury JF., Robinson M.R., Marchini J.L., and Dermitzakis E.T. (2019) Accurate, scalable and integrative haplotype estimation. Nat Commun.10(1):5436. doi.org/10.1038/s41467-019-13225-y

Gualdrón Duarte J.L., Gori AS., Hubin X., Lourenco D., Charlier C. *et al.* (2020) Performances of Adaptive MultiBLUP, Bayesian regressions, and weighted-GBLUP approaches for genomic predictions in Belgian Blue beef cattle. BMC Genomics 21, 545. doi.org/10.1186/s12864-020-06921-3

Howie B., Fuchsberger C., Stephens M., Marchini J., Abecasis GR. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 44(8): 955–959. doi:10.1038/ng.2354.

Stephens M. (2013) A Unified Framework for Association Analysis with Multiple Related Phenotypes. PLoS ONE 8(7): e65245. doi:10.1371/journal.pone.0065245

Zhou X., and Stephens M. (2012) Genome-wide efficient mixed-model analysis for association studies. Nat Genet 44(7): 821–824. doi:10.1038/ng.2310