



A Meta-Analysis Including Pre-selected Sequence Variants Associated With Seven Traits in Three French Dairy Cattle Populations

Andrew G. Marete^{1,2*}, Bernt Gulbrandsen², Mogens S. Lund², Sébastien Fritz^{1,3}, Goutam Sahana² and Didier Boichard¹

¹ UMR GABI, INRA, AgroParisTech, Université Paris Saclay, 78350 Jouy en Josas, France, ² Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark, ³ ALLICE, Paris, France

OPEN ACCESS

Edited by:

Tad Stewart Sonstegard,
Recombinetics, United States

Reviewed by:

Dan Nonneman,
Agricultural Research Service (USDA),
United States
Luca Fontanesi,
Università degli Studi di Bologna, Italy

*Correspondence:

Andrew G. Marete
didier.boichard@inra.fr

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 03 June 2018

Accepted: 16 October 2018

Published: 06 November 2018

Citation:

Marete AG, Gulbrandsen B, Lund MS, Fritz S, Sahana G and Boichard D (2018) A Meta-Analysis Including Pre-selected Sequence Variants Associated With Seven Traits in Three French Dairy Cattle Populations. *Front. Genet.* 9:522. doi: 10.3389/fgene.2018.00522

A within-breed genome-wide association study (GWAS) is useful when identifying the QTL that segregates in a breed. However, an across-breed meta-analysis can be used to increase the power of identification and precise localization of QTL that segregate in multiple breeds. Precise localization will allow including QTL information from other breeds in genomic prediction due to the persistence of the linkage phase between the causal variant and the marker. This study aimed to identify and confirm QTL detected in within-breed GWAS through a meta-analysis in three French dairy cattle breeds. A set of sequence variants selected based on their functional annotations were imputed into 50 k genotypes for 46,732 Holstein, 20,096 Montbeliarde, and 11,944 Normande cows to identify QTL for milk production, the success rate at insemination of cows (fertility) and stature. We conducted within-breed GWAS followed by across-breed meta-analysis using a weighted Z-scores model on the GWAS summary data (i.e., *P*-values, effect direction, and sample size). After Bonferroni correction, the GWAS result identified 21,956 significantly associated SNP ($P_{\text{FWER}} < 0.05$), while meta-analysis result identified 9,604 significant SNP ($P_{\text{FWER}} < 0.05$) associated with the phenotypes. The meta-analysis identified 36 QTL for milk yield, 48 QTL for fat yield and percentage, 29 QTL for protein yield and percentage, 13 QTL for fertility, and 16 QTL for stature. Some of these QTL were not significant in the within-breed GWAS. Some previously identified causal variants were confirmed, e.g., BTA14:1802265 (fat percentage, $P = 1.5 \times 10^{-760}$; protein percentage, $P = 7.61 \times 10^{-348}$) both mapping the DGAT1-K232A mutation and BTA14:25006125 ($P = 8.58 \times 10^{-140}$) mapping *PLAG1* gene was confirmed for stature in Montbeliarde. New QTL lead SNP shared between breeds included the intronic variant rs109205829 (*NFIB* gene), and the intergenic variant rs41592357 (1.38 Mb upstream of the *CNTN6* gene and 0.65 Mb downstream of the *CNTN4* gene). Rs110425867 (*ZFAT* gene) was the top variant associated with fertility, and new QTL lead SNP included rs109483390 (0.1 Mb upstream of the *TNFAIP3* gene and 0.07 Mb downstream of *PERP* gene), and rs42412333 (0.45 Mb downstream of the *RPL10L* gene). An across-breed meta-analysis had greater power to detect QTL as opposed to a within breed GWAS. The QTL detected here can be incorporated in routine genomic predictions.

Keywords: candidate genes, candidate SNP, GWAS, meta-analysis, sequence imputation

INTRODUCTION

Ideally, the magnitude of estimated effects on the quantitative trait of interest could be used to rank single nucleotide polymorphisms (SNP) for a functional genomic study to identify causal variants. The physical locations of the associated SNP on the genome are then flagged and mined for causative variants underlying quantitative trait loci (QTL). Causal variants reported for QTL in previous dairy cattle studies include polymorphisms causing variation in milk production, fertility, and embryo mortality (Grisart et al., 2002; Hoff et al., 2017; Michot et al., 2017; Bouwman et al., 2018). Most analyses testing genomic regions, like those above, perform within-breed GWAS (Chang et al., 2018). Even so, our knowledge regarding causal variants in these genomic regions remains limited because the GWAS analyses yield similar *P*-values for many adjacent SNP variants, a consequence of linkage disequilibrium (LD). Even with a multitude of GWAS results, strong LD prevents distinguishing the actual causal variant from linked markers (Goddard and Hayes, 2009). A meta-analysis can be used to improve the resolution of QTL detection and identify causal variants provided that LD is conserved at short distances across breeds (van den Berg et al., 2016). The main advantage of a meta-analysis is that it allows simultaneous analysis of many breeds by combining GWAS summary statistics across populations, thereby increasing power to detect QTL (van den Berg et al., 2016; Bouwman et al., 2018).

In this paper, we report an across-breed meta-analysis of GWAS summary statistics based on imputed pre-selected sequence variant genotypes from 78,772 cattle from three dairy breeds and for seven traits. The meta-analysis across breeds allowed us to identify 142 QTL for milk production, stature, and fertility.

MATERIALS AND METHODS

Studied Population and Phenotypes

We studied phenotypes from three French dairy cattle breeds. No ethical approval was necessary since the data used was from existing databases. Phenotypes were from cows born between 2007 and 2013 in the French dairy production regions and were defined as follows: (1) Production traits including milk, protein, and fat yields were obtained from test-day records expressed as 305 d yield (kg). Fat content (g/kg) was calculated as $1,000 \times (\text{fat yield}/\text{milk yield})$. Protein content (g/kg) was calculated as $1,000 \times (\text{protein yield}/\text{milk yield})$. (2) Fertility was defined as success/failure at each insemination of lactating cows and recorded as a success (=1) or a failure (=0). (3) Stature was measured as the vertical distance from the plank to the sacrum.

These traits were analyzed using the French national evaluation model (Boichard et al., 2012b) to obtain Yield Deviations (YD) (VanRaden and Wiggans, 1991). A YD can be defined as the mean performance adjusted for all environmental effects, including the permanent environment for cows with multiple records. The model varied according to the trait, especially concerning environmental factors. We retained traits

for genotyped cows and this included data from 46,732 Holstein cows, 20,096 Montbeliarde cows, and 11,944 Normande cows.

Genotyping and Imputation

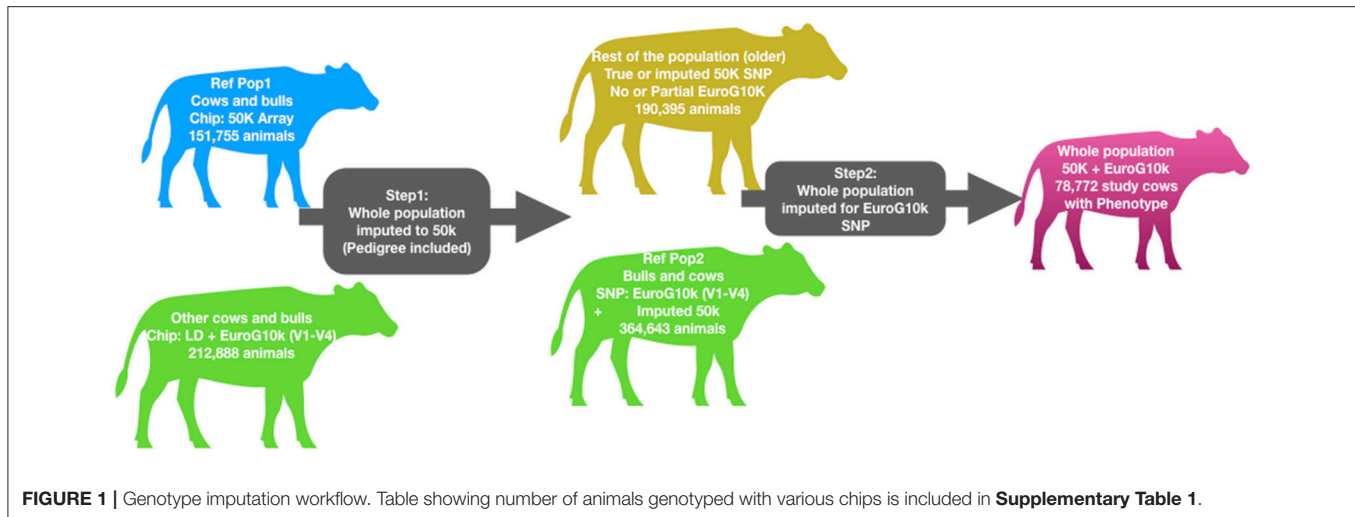
Cows were genotyped using the Illumina Infinium® BovineSNP50 BeadChip (50 K, Illumina, San Diego, CA), BovineLD BeadChip (Boichard et al., 2012a) or one of the first four versions of the EuroG10k SNP chip (Boichard et al., 2018). The EuroG10k SNP chip is composed of two parts: (1) Between 7,000 and 8,500 generic (and supposedly neutral) SNP from BovineLD Genotyping BeadChip v.2 (Boichard et al., 2012a). (2) Whole genome sequence variants—a custom part of up to 7,232 SNP selected from sequence data as part of 1,000 Bull Genomes Project *Run 4* (Daetwyler et al., 2014) based on their functional annotation. The full description of the EuroG10k chip and quality control procedure was previously described by Marete et al. (2018a) and Boichard et al. (2018). Cows with phenotypes were rather old animals and were genotyped with the 50 k chip. Younger animals including cows without phenotypes were genotyped with the EuroG10k chip. To get complete marker information across animals, and to include the sequence variants in the analysis, we run two successive steps of imputation for the entire populations (males and females, with and without phenotypes, see **Figure 1**). First, we imputed all 50 k markers for all animals using animals with 50 k genotypes as a reference. Then we imputed sequence variants using the animals genotyped with the EuroG10k chip as the reference population. The imputation was done within breed using the FImpute software (Sargolzaei et al., 2014). After imputation, 48,576 SNP distributed across 29 *Bos taurus* autosomes (BTA) remained. Of these, 42,967 were from the 50 k chip, and 5,609 sequence variants selected from French cattle populations. Imputation accuracy was estimated using both concordance rate and allelic squared correlation (r^2), both for animals genotyped with 50 k only and those genotyped with EuroG10k chip only. In both instances, we randomly masked 20% of the genotyped markers on each chromosome and 15% of the cows per breed, then performed an imputation and compared the imputed genotypes with true genotypes by estimating a Pearson correlation coefficient. For the 50 k SNP, the average concordance was >0.98 , and average allelic r^2 was >0.97 , and for the sequence variants, the average concordance was >0.95 , and average r^2 was >0.96 for all breeds.

Association Studies, Meta-Analysis, and QTL Heterogeneity

First, 48,576 SNP (pre-selected sequence variants and 50 k SNP) were tested individually for association with each trait within the breed. The GCTA software (Yang et al., 2011) was used to fit a mixed linear regression model to test associations between a SNP and the trait. For any given trait, the fitted model was

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{x}\beta + \mathbf{g} + \varepsilon \quad (1)$$

where \mathbf{y} was a vector of phenotypes (yield deviation) for all cows, $\mathbf{1}$ was a row vector of 1s, μ was the mean, \mathbf{x} was the vector of true or imputed allele dosages for the SNP, β was the



fixed allele substitution effect of the SNP, \mathbf{g} was the vector of random additive genetic effects with $\mathbf{g} \sim N(0, \mathbf{G}\sigma_g^2)$, \mathbf{G} being the genomic relationship matrix, σ_g^2 was the variance explained by all the SNPs, $\boldsymbol{\varepsilon}$ was a vector of random residual effects with $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\sigma_\varepsilon^2)$, σ_ε^2 was the error variance, and \mathbf{I} was an identity matrix. The variance of \mathbf{y} was $\text{var}(\mathbf{y}) = \mathbf{G}\sigma_g^2 + \mathbf{I}\sigma_\varepsilon^2$. The g_{jk} term of \mathbf{G} matrix was estimated by

$$g_{jk} = \frac{1}{w} \sum_{i=1}^w \frac{(z_{ij} - 2p_i)(z_{ik} - 2p_i)}{2p_i(1 - p_i)} \quad (2)$$

where w was the total number of SNP, z_{ij} and z_{ik} were numbers of copies of the reference allele for the i^{th} SNP in the j^{th} and k^{th} cows, respectively, and p_i was the frequency of the reference allele estimated from the marker data (VanRaden, 2008).

To control the type I error rate and assuming all tests were independent (a conservative assumption due to LD among SNP), a Bonferroni genome-wide correction was applied for the total number of tests for each trait i.e., the nominal type I error rate ($\alpha = 5\%$) was divided by the number of SNP ($w = 48,576$) to obtain a genome-wide error rate threshold of $t = 1.03 \times 10^{-6}$. Any SNP whose probability of observing the test statistics was below this value was considered genome-wide significant.

The within-breed analyses were followed by a meta-analysis using weighted Z-score model as implemented in METAL software (Willer et al., 2010). The weighted Z-score model used P -values and directions of effect estimates and weights individual GWAS based on the sample size to compute a Z-score, i.e.,

$$Z_b = \Phi^{-1} \left(1 - \frac{P_b}{2} \right) * \Delta b \quad (3)$$

where Z_b was the Z-score for breed b , Φ was the standard normal cumulative distribution, P_b was the P -value, Δb was the direction of the SNP effect estimated within breed b . Overall Z-score was calculated as:

$$Z = \frac{\sum_b Z_b w_b}{\sqrt{\sum_b w_b^2}} \quad (4)$$

where $w_b = \sqrt{N_b}$ and N_b was the sample size for breed b . The overall P -value was calculated as:

$$P = 2\Phi(-|Z|) \quad (5)$$

We evaluated heterogeneity of the effect sizes across breeds using Cochran's Q-test (Cochran, 1954) as implemented in the METAL software. A QTL was deemed to segregate across populations if the heterogeneity P -value was ≤ 0.05 . The functional consequence of significantly associated variants was predicted using the Variant Effect Predictor tool from Ensembl Genome Browser 90 (McLaren et al., 2016). We classified variants that (i) mapped to a gene, (ii) <5 -kb to known genes, and, (iii) >5 -kb from any coding region. In case of multiple variants representing the same gene, we kept the variant associated with most traits and with the lowest P -value as representative for that gene.

Identification of QTL That Segregates Within and Across Breeds

Since it is challenging to identify the causative variant, we aimed at determining the QTL region that most likely harbor the causal variant. First, for each chromosome, we subjectively inspected significantly associated variants within the genomic regions. Secondly, we defined the within-breed QTL region as 1 Mb windows and the SNP with the lowest P -value designated as the lead SNP. A QTL was selected as significant in more than one breed if any of the SNP in the QTL region of the other breed had a P -value lower than the genome-wide threshold ($P < 1.03 \times 10^{-6}$). The UMD3.1 (Zimin et al., 2009) bovine genome annotation was used to annotate genes centering within 1 Mb intervals on the lead SNP. We compared these annotations to known QTL for bovine milk production, fertility and stature traits using QTLdb (Hu et al., 2016) and literature review.

RESULTS

Association Studies

Overall, there were 21,956 SNP significantly associated with traits in three French cattle breeds. Forty-three percent of these associated SNP were from Holstein, 24% from Montbeliarde, 14% from Normande. Seven percent of these SNP showed association in both Holstein and Montbeliarde, 4% in Holstein and Normande, 5% in Montbeliarde and Holstein, and 3% in all three breeds. Production traits and stature had more significant hits across breeds, compared to fertility. As presented in **Figures 2A–E**, the overlap of significant tests in any two-breed combination and for all traits was most evident between Holstein and Montbeliarde (1,621 SNP) and least evident between Montbeliarde and Normande (1,082 SNP). Seven hundred and one SNP were significantly associated with the same trait in all three breeds.

As presented in **Figures 2F–I**, some of the imputed sequence variants had lower P -values compared with the 50k variants P -values. Although more than 13% of the pre-selected sequence variants had MAF between 0.5 and 5%, only 2 QTL lead SNP had MAF lower than 5%. We did not detect QTL with MAF <5% in Normande. Among the pre-selected sequence variants, 1,478 were significant ($P < 1.03 \times 10^{-6}$).

Within Breed GWAS And Multi-Breed Meta-Analysis Production Traits

Milk yield

We observed both known and novel QTL in all three cattle breeds. The most significant QTL lead SNP from meta-analysis are presented in **Table 1**. We observed QTL on 16 chromosomes where Holstein had 23 QTL, Montbeliarde had 10 QTL, Normande had 15 QTL. The most significant lead SNP was located at *Bos taurus autosome* (BTA) 14:1801116 (rs109421300, $P = 9.7 \times 10^{-283}$), an intronic variant near the causal variants in the *DGAT1* gene. Other significant QTL were observed at BTA5:50804085 ($P = 1.3 \times 10^{-22}$, *intergenic*), rs109205829 (BTA8:30088133, $P = 2.3 \times 10^{-21}$, *NFIB* gene), and BTA20:32670639 ($P = 4.0 \times 10^{-41}$, *intergenic* flanking 3' end of the *GHR* gene).

A QTL on BTA8 at 29–30 Mb with the same lead SNP at BTA8:30088133 (rs109205829) was observed in both Holstein and Montbeliarde. This lead SNP is an intronic variant within the *NFIB* gene, and the meta-analysis had a lower P -value ($P = 2.3 \times 10^{-22}$) than either of the single breed GWAS ($P = 8.4 \times 10^{-14}$). Another QTL on BTA22 at 23–26 Mb with a lead SNP at BTA22:24219999 (rs41592357) was observed in all three breeds. Rs41592357 is an intergenic variant 1.38 Mb upstream of the *CNTN6* gene and 0.65 Mb downstream of the *CNTN4* gene.

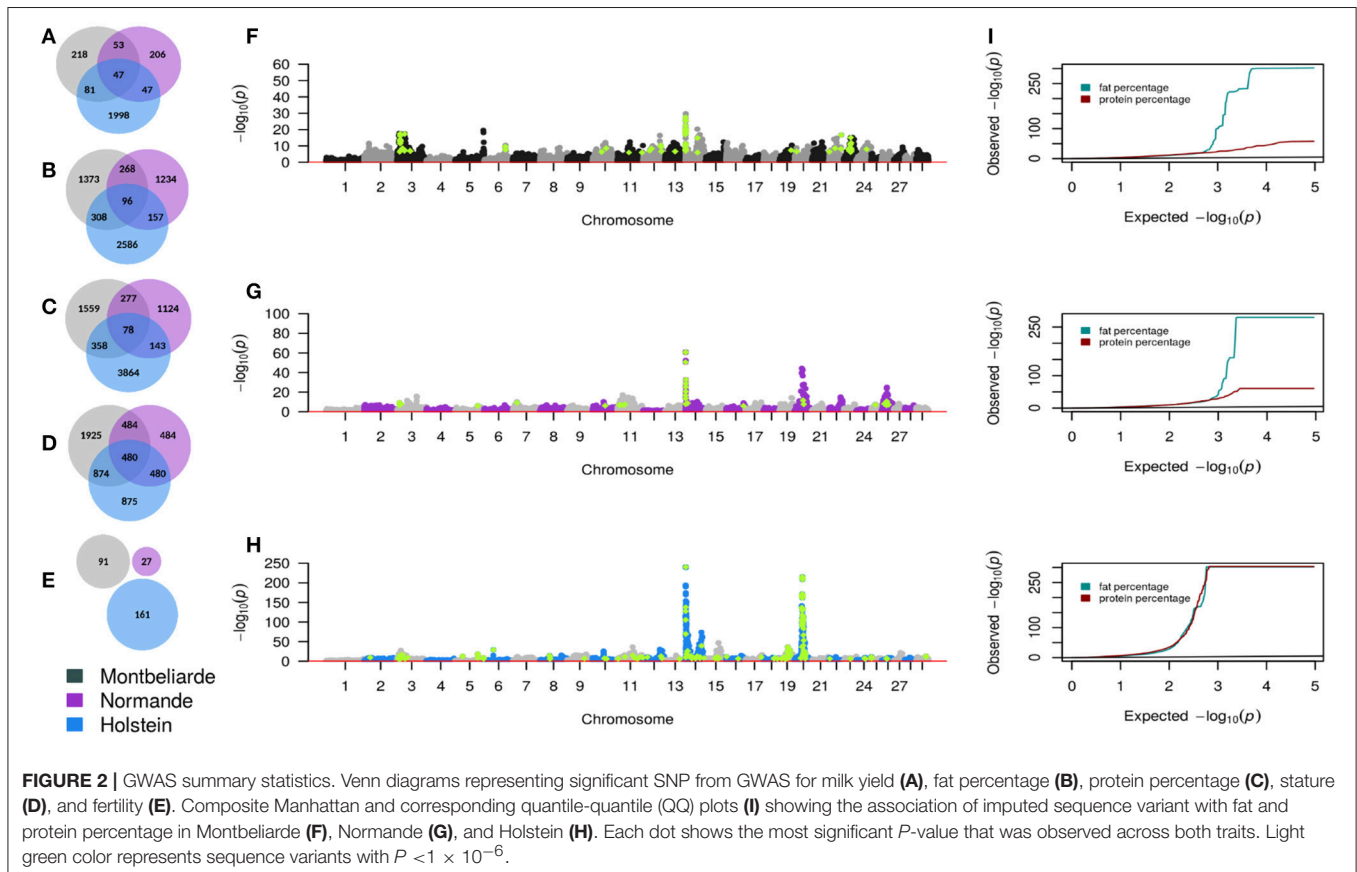


TABLE 2 | Continued

| QTL | Reference SNP cluster ID | CHR | BP | Allele | Nearby gene | P _{meta} | Effect direction (and evidence of heterogeneity) across three populations ¹ | MON | | | NOR | | | HOL | | |
|-----|--------------------------|-----|----------|--------|--------------------|-------------------|--|-------------|----------|-------------|----------|-------------|-----------|-------|---|--|
| | | | | | | | | b(se) | p | b(se) | p | b(se) | p | b(se) | p | |
| 36 | rs109882115 | 18 | 58067310 | a/g | ENSBTAG00000039491 | 9.00e-28 | +++(7.0e-03) | 0.20(0.04) | 1.50e-07 | 0.06(0.04) | 1.30e-01 | 0.24(0.02) | 2.20e-23 | | | |
| 37 | rs109751680 | 19 | 42976859 | a/g | STAT5B | 1.60e-37 | ++(-2.5e-18) | 0.08(0.03) | 4.80e-03 | -0.04(0.04) | 2.50e-01 | 0.34(0.02) | 3.50e-53 | | | |
| 38 | rs41921177 | 19 | 51326750 | a/g | CCDC57 | 6.80e-25 | --(-1.2e-03) | 5.28(0.87) | 1.40e-09 | 0.58(0.91) | 5.30e-01 | 2.69(0.30) | 1.00e-19 | | | |
| 39 | rs207867588 | 20 | 31128620 | t/c | ENSBTAG00000033187 | 6.60e-166 | ?+?(1.0e+00) | -0.64(0.55) | 2.40e-01 | -0.01(0.04) | 8.80e-01 | -0.84(0.03) | 6.60e-166 | | | |
| 40 | rs109262355 | 20 | 35249040 | t/c | FYB | 3.50e-16 | --(-1.3e-04) | -0.36(0.48) | 4.50e-01 | -1.50(0.58) | 9.50e-03 | -2.56(0.29) | 1.60e-18 | | | |
| 41 | rs137102117 | 23 | 27450989 | a/g | LY6G5B | 4.20e-20 | ?+?(1.0e+00) | 0.38(0.05) | 7.60e-15 | 0.59(0.06) | 4.20e-20 | 0.14(0.24) | 5.60e-01 | | | |
| 42 | rs110212032 | 24 | 37652964 | a/g | - | 6.40e-19 | +++(8.0e-04) | 0.10(0.02) | 3.80e-05 | 0.01(0.03) | 8.40e-01 | 0.18(0.02) | 2.60e-18 | | | |
| 43 | rs41659834 | 26 | 14814944 | t/c | MYOF | 1.30e-17 | +++(4.3e-01) | -1.36(0.40) | 6.30e-04 | -1.99(0.47) | 2.10e-05 | -2.05(0.31) | 2.10e-11 | | | |
| 44 | rs42090237 | 26 | 20290497 | t/c | GOT1 | 3.00e-19 | +++(2.0e-01) | 0.08(0.03) | 1.90e-03 | 0.24(0.05) | 6.60e-06 | 0.15(0.02) | 2.40e-13 | | | |
| 45 | rs109330072 | 26 | 32442734 | t/g | - | 5.10e-13 | --(-1.7e-01) | -4.00(1.01) | 7.50e-05 | -5.09(4.75) | 2.80e-01 | -2.54(0.41) | 4.40e-10 | | | |
| 46 | rs109877780 | 27 | 26628005 | t/g | - | 1.90e-35 | +++(4.7e-04) | -0.10(0.03) | 9.30e-05 | -0.13(0.04) | 3.00e-03 | -0.25(0.02) | 1.50e-38 | | | |
| 47 | rs109599512 | 27 | 36117365 | a/g | GOLGA7 | 2.00e-72 | --(-1.7e-03) | 0.21(0.03) | 9.30e-16 | 0.39(0.04) | 8.70e-25 | 0.29(0.02) | 4.50e-33 | | | |
| 48 | rs42181608 | 29 | 29254185 | a/g | PKNOX2 | 3.70e-13 | --(-3.9e-02) | -0.62(0.41) | 1.30e-01 | -1.80(0.46) | 1.10e-04 | -1.95(0.30) | 9.10e-11 | | | |

¹Populations in order are MON, NOR, and HOL; "+", "++" and "-" denote positive and negative substitution effects of the alternate allele. "?" indicates that the variant did not segregate in the respective population. The P-value of Cochran's Q-test for heterogeneity of the effect sizes across breeds is given in parentheses, and is significant if P < 0.05.

Some QTL overlapped between Holstein and Montbeliarde, some with same lead SNP, but others with different lead SNP, e.g., BTA15 at 57–61 Mb and BTA25 at 10–11 Mb. The lead SNP for the QTL on BTA15 were rs110049689 (BTA15:57333896, $P = 1.5 \times 10^{-7}$, *MYO7A* gene) in Montbeliarde, and rs41690133 (BTA15:56842162, $P = 2.4 \times 10^{-12}$, intergenic) in Holstein. On BTA25, the lead SNP was rs42062121 (BTA25:11015593, *SNX29* gene) in both Holstein ($P = 1.1 \times 10^{-7}$) and Montbeliarde ($P = 2.5 \times 10^{-6}$).

Fat yield and percentage.

We observed QTL on 19 chromosomes where Holstein had 26 QTL, Montbeliarde had 12 QTL, Normande had 5 QTL. Meta-analysis confirmed BTA14:1802265 ($P = 1.5 \times 10^{-760}$, *DGAT1* gene). Other QTL lead SNP whose meta-analysis P-values exceeded the GWAS significance threshold include: rs110824611 (BTA3:11040167, $P = 1.2 \times 10^{-7}$, *OR6K6* gene), rs41616289 (BTA8:93896073, $P = 2.9 \times 10^{-7}$, intergenic), rs109394729 (BTA10:46486647, $P = 1.9 \times 10^{-7}$, *USP3* gene), rs41668653 (BTA11:63467507, $P = 1.0 \times 10^{-10}$, intergenic), rs110249976 (BTA15:53166998, $P = 4.8 \times 10^{-9}$, *FCHSD2* gene), rs41921177 (BTA19:51326750, $P = 8.0 \times 10^{-14}$, *CCDC57* gene), and rs109599512 (BTA27:36117365, $P = 9.4 \times 10^{-17}$, *GOLGA7* gene).

We observed QTL overlapping in Holstein and Montbeliarde and in Holstein and Normande, but with different lead SNP. The two lead SNP for QTL significantly associated in both Holstein and Montbeliarde were BTA12:15173534 ($P = 3.1 \times 10^{-13}$, *NUFIP1* gene, 14.5–15.6 Mb), and rs41921177 (BTA19:51326750, $P = 1.1 \times 10^{-19}$, *CCDC57* gene, 51–52 Mb). The three lead SNP for QTL significantly associated in both Holstein and Normande were rs109394729 (BTA10:46486647, *USP3* gene, 43–47 Mb), rs109599512 (BTA27:36117365, *GOLGA7* gene, and 35.5–38.5 Mb), and rs41668653 (BTA11:63467507, intergenic, 61.2–64.9 Mb).

Protein yield and percentage.

We observed QTL on 22 chromosomes where Holstein had 31 QTL, Montbeliarde had 18 QTL, Normande had 22 QTL. We confirmed already known causal variants, BTA6:87199843 ($P = 9.5 \times 10^{-31}$, *CSN2* gene), and BTA14:1802265 ($P = 8.0 \times 10^{-78}$, *DGAT1* gene). As presented in Table 4, new QTL lead SNP whose meta-analysis P-value exceeded the within breed GWAS significance levels included: rs41593345 (BTA9:45303296, $P = 4.3 \times 10^{-18}$, *PREP* gene), rs41668653 (BTA11:63467507, $P = 7.5 \times 10^{-48}$, *RABIA* gene), and rs109882115 (BTA18:58067310, $P = 9.0 \times 10^{-28}$, *CEACAM18* gene). Two QTL overlaps were observed between Holstein and Montbeliarde, on BTA6 (85–88 Mb) with the most significant lead SNP at BTA6:87199843 (*HSTN*), and on BTA21 with lead SNP. rs29011638 (BTA21:37684315, $P = 3.8 \times 10^{-7}$, intergenic). We observed two QTL overlapping between Holstein and Normande on BTA7 (13–14 Mb) with the most significant SNP, rs41568613 (BTA7:13526016, $P = 7.7 \times 10^{-6}$, intergenic), and on BTA26 (21.5–24.5 Mb) with lead SNP, rs29014382 (BTA26:22339074, $P = 1.8 \times 10^{-9}$, intergenic). We observed two QTL overlapping among all three breeds on BTA16 (1–2 Mb) with lead SNP, rs42450080 (BTA16:1608132, $P = 1.2$

TABLE 3 | Lead SNP variants at 29 QTL for protein yield and percentage in milk.

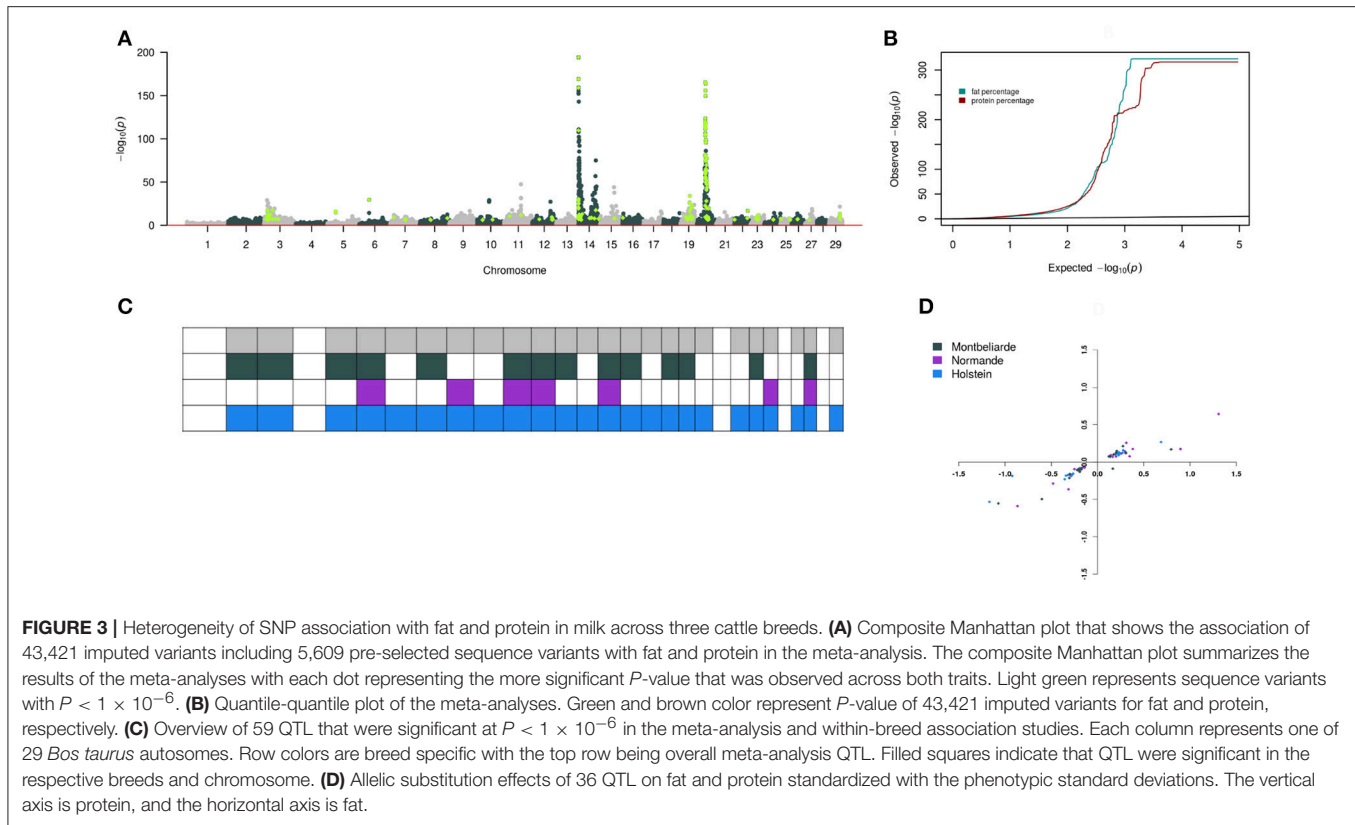
| QTL | Reference SNP cluster ID | CHR | BP | Allele | Nearby gene | P _{meta} | Effect direction (and evidence of heterogeneity) across three populations ¹ | MON | | | NOR | | | HOL | | |
|-----|--------------------------|-----|-----------|--------|-------------|-------------------|--|-------------|---------|-------------|---------|-------------|---------|-------|---|--|
| | | | | | | | | b(se) | P | b(se) | P | b(se) | P | b(se) | P | |
| 1 | rs29025329 | 2 | 44947955 | a/g | NMI | 2.94e-18 | +++(9.6e-02) | -0.82(0.32) | 1.0e-02 | -1.28(0.36) | 3.6e-04 | -1.61(0.21) | 4.8e-15 | | | |
| 2 | rs42889305 | 2 | 70680985 | t/g | — | 1.15e-18 | --(3.2e-04) | 0.36(0.33) | 2.8e-01 | 1.52(0.43) | 4.5e-04 | 1.84(0.20) | 3.0e-19 | | | |
| 3 | rs110532777 | 3 | 21640254 | t/c | POLR3C | 8.05e-11 | --(3.8e-02) | 0.49(0.30) | 1.0e-01 | 0.52(0.35) | 1.5e-01 | 1.38(0.21) | 3.3e-11 | | | |
| 4 | rs208464932 | 3 | 48555688 | a/g | SLC44A3 | 1.17e-13 | ??+(1.0e+00) | — | — | — | — | 2.02(0.27) | 1.2e-13 | | | |
| 5 | rs42737927 | 3 | 113907407 | t/c | UGT1A1 | 3.31e-12 | +++(4.3e-05) | -0.49(0.34) | 1.4e-01 | 0.05(0.35) | 8.9e-01 | -1.73(0.21) | 3.6e-16 | | | |
| 6 | rs43703011 | 6 | 87181619 | t/g | CSN2 | 9.84e-31 | --(6.4e-03) | -1.92(0.42) | 5.9e-06 | -3.71(0.51) | 2.2e-13 | -2.67(0.32) | 1.2e-16 | | | |
| 7 | rs41609745 | 7 | 65358446 | t/c | — | 1.89e-14 | +++(2.4e-04) | 0.11(0.32) | 7.3e-01 | 1.52(0.39) | 8.6e-05 | 1.57(0.20) | 1.0e-14 | | | |
| 8 | rs109205829 | 8 | 30088133 | c/g | NFIB | 1.79e-17 | +++(3.2e-01) | 1.22(0.32) | 1.2e-04 | 1.69(0.36) | 2.7e-06 | 1.29(0.21) | 7.4e-10 | | | |
| 9 | rs42254183 | 8 | 65107050 | a/g | — | 5.25e-12 | +++(1.2e-03) | -0.26(0.43) | 5.5e-01 | -1.13(0.53) | 3.4e-02 | -2.20(0.29) | 6.7e-14 | | | |
| 10 | rs109630763 | 8 | 103526477 | t/c | — | 1.85e-12 | --(1.5e-03) | -5.37(1.26) | 2.0e-05 | 0.20(0.35) | 5.6e-01 | -1.79(0.27) | 3.0e-11 | | | |
| 11 | rs41612115 | 9 | 51985736 | t/c | — | 1.64e-22 | +++(4.8e-02) | 1.32(0.37) | 3.8e-04 | 1.19(0.48) | 1.3e-02 | 2.46(0.27) | 1.1e-19 | | | |
| 12 | rs41647455 | 10 | 15333419 | a/g | — | 1.22e-16 | --(1.1e-07) | -0.32(0.46) | 4.9e-01 | -0.23(0.38) | 5.3e-01 | -2.12(0.21) | 1.8e-23 | | | |
| 13 | rs43625971 | 10 | 47689396 | t/c | TLN2 | 1.47e-13 | +++(9.7e-03) | -0.36(0.32) | 2.7e-01 | -1.25(0.35) | 3.7e-04 | -1.70(0.24) | 1.5e-12 | | | |
| 14 | rs109186903 | 10 | 94083525 | t/c | — | 2.51e-18 | ??+(1.0e+00) | — | — | — | — | 2.23(0.26) | 2.5e-18 | | | |
| 15 | rs41684057 | 13 | 17104043 | t/c | PRKCG | 1.48e-10 | +++(8.7e-02) | 1.14(0.83) | 1.7e-01 | 1.19(0.47) | 1.1e-02 | 1.30(0.21) | 8.6e-10 | | | |
| 16 | rs110991150 | 13 | 40815971 | c/g | — | 3.29e-20 | +++(4.0e-02) | 1.59(0.46) | 5.1e-04 | 7.27(3.66) | 4.7e-02 | 2.64(0.30) | 4.1e-18 | | | |
| 17 | rs109421300 | 14 | 18011116 | t/c | HSF1 | 8.03e-78 | +++(8.2e-11) | -8.08(1.59) | 3.7e-07 | -2.70(0.62) | 1.5e-05 | -4.91(0.26) | 3.1e-78 | | | |
| 18 | rs41617796 | 14 | 37670101 | t/c | — | 3.11e-11 | +++(1.6e-01) | 0.63(0.31) | 4.0e-02 | 0.74(0.40) | 6.2e-02 | 1.71(0.27) | 2.4e-10 | | | |
| 19 | rs109020671 | 14 | 60893281 | a/g | — | 2.25e-12 | +++(4.4e-01) | -0.81(0.33) | 1.5e-02 | -1.25(0.43) | 3.8e-03 | -1.29(0.21) | 1.5e-09 | | | |
| 20 | rs42221575 | 15 | 60250434 | a/g | — | 2.21e-14 | +++(1.4e-02) | -0.63(0.46) | 1.7e-01 | -1.28(0.39) | 1.1e-03 | -1.49(0.20) | 1.7e-13 | | | |
| 21 | rs109053739 | 15 | 77590401 | g/c,t | CKAP5 | 6.60e-13 | +++(3.1e-03) | -2.24(0.64) | 4.6e-04 | 0.12(0.63) | 8.4e-01 | -1.93(0.27) | 9.7e-13 | | | |
| 22 | rs41628722 | 17 | 60612635 | a/c | — | 1.48e-16 | --(2.8e-03) | 0.94(0.31) | 2.7e-03 | 0.26(0.36) | 4.7e-01 | 2.12(0.25) | 4.7e-17 | | | |
| 23 | rs110031071 | 19 | 51767413 | a/g | PDEF6G | 1.59e-09 | --(5.7e-02) | -0.36(0.31) | 2.3e-01 | -1.34(0.36) | 1.6e-04 | -1.06(0.21) | 2.6e-07 | | | |
| 24 | rs436825729 | 20 | 32670639 | a/g | — | 3.66e-304 | ??+(1.0e+00) | — | — | — | — | -0.59(0.01) | 0.00 | | | |
| 25 | rs41592357 | 22 | 24219999 | t/c | — | 5.63e-16 | --(4.3e-01) | -1.22(0.35) | 4.6e-04 | -1.50(0.35) | 1.8e-05 | -1.24(0.20) | 1.5e-09 | | | |
| 26 | rs385583887 | 23 | 29349996 | a/g | — | 1.20e-17 | +?(1.2e-07) | 0.08(0.31) | 7.9e-01 | — | — | 2.81(0.28) | 9.0e-24 | | | |
| 27 | rs41624917 | 26 | 15383866 | a/g | PLCE1 | 2.09e-11 | +++(5.0e-03) | 0.87(0.39) | 2.6e-02 | 0.14(0.51) | 7.9e-01 | 1.46(0.21) | 1.2e-12 | | | |
| 28 | rs109839618 | 26 | 42403875 | a/g | TACC2 | 8.56e-13 | +++(1.6e-02) | -0.64(0.31) | 3.9e-02 | -0.47(0.40) | 2.4e-01 | -1.91(0.26) | 2.2e-13 | | | |
| 29 | rs41686102 | 29 | 41989397 | t/c | SLC22A6 | 1.38e-51 | +++(1.1e-16) | -0.03(0.02) | 3.3e-02 | -0.06(0.03) | 3.0e-02 | -2.3(0.01) | 1.4e-65 | | | |

¹Populations in order are MON, NOR, and HOL; “+” and “-” denote positive and negative substitution effects of the alternate allele. “?” indicates that the variant did not segregate in the respective population. The P-value of Cochran’s Q-test for heterogeneity of the effect sizes across breeds is given in parentheses, and is significant if P < 0.05.

TABLE 5 | Lead SNP variants at 16 QTL for stature.

| QTL | Reference SNP cluster ID | CHR | BP | Allele | Nearby gene | P _{meta} | Effect direction (and evidence of heterogeneity) across three populations ¹ | MON | | | NOR | | | HOL | | |
|-----|--------------------------|-----|-----------|--------|-------------|-------------------|--|-------------|----------|-------------|---------|-------------|---------|-------|---|---|
| | | | | | | | | b(se) | p | P | b(se) | p | P | b(se) | p | P |
| 1 | rs41645172 | 2 | 109778246 | t/c | — | 3.3e-10 | +++(8.1e-10) | 0.31(0.04) | 3.9e-17 | 0.22(0.07) | 2.3e-03 | 0.18(0.01) | 2.7e-01 | | | |
| 2 | rs43121344 | 7 | 68281468 | t/c | — | 7.6e-19 | --(4.6e-01) | 0.30(0.05) | 4.5e-08 | 0.44(0.12) | 3.0e-05 | 0.04(0.01) | 1.1e-09 | | | |
| 3 | rs110652403 | 7 | 90460692 | t/c | — | 1.2e-11 | --(1.3e-02) | -0.04(0.04) | 3.0e-01 | -0.26(0.01) | 1.5e-02 | -0.05(0.01) | 4.8e-12 | | | |
| 4 | rs109703572 | 9 | 46739973 | t/c | — | 1.8e-23 | ++++(6.3e-03) | -0.09(0.04) | 1.8e-02 | -0.22(0.06) | 7.7e-05 | -0.06(0.01) | 4.9e-21 | | | |
| 5 | rs41664341 | 9 | 80859295 | a/g | — | 5.8e-22 | --(2.6e-02) | 0.15(0.04) | 5.2e-05 | 0.10(0.06) | 8.7e-02 | 0.06(0.01) | 2.5e-19 | | | |
| 6 | rs109860141 | 10 | 51558941 | a/g | FAM63B | 5.9e-16 | ++++(2.6e-09) | 0.34(0.04) | 2.4e-20 | 0.22(0.06) | 4.2e-04 | 0.02(0.01) | 7.9e-03 | | | |
| 7 | rs110128058 | 10 | 58318595 | a/g | GNB5 | 3.5e-18 | ++++(3.5e-09) | 0.40(0.04) | 3.6e-22 | 0.22(0.07) | 2.5e-03 | 0.04(0.01) | 6.4e-04 | | | |
| 8 | rs43430263 | 12 | 53241975 | t/g | SLAIN1 | 6.5e-16 | -(1.4e-04) | 0.19(0.04) | 5.7e-07 | -0.04(0.06) | 4.6e-01 | 0.06(0.01) | 3.3e-14 | | | |
| 9 | rs29019889 | 13 | 33366445 | a/g | NSUN6 | 6.4e-16 | ++++(9.0e-03) | 0.21(0.03) | 4.5e-10 | 0.10(0.10) | 3.3e-01 | 0.04(0.01) | 3.5e-09 | | | |
| 10 | rs134215421 | 14 | 25006125 | t/c | PLAG1 | 8.6e-140 | ??(1.0e+00) | 1.70(0.07) | 8.6e-140 | — | — | — | — | | | |
| 11 | rs41584869 | 19 | 17569423 | a/g | SPACA3 | 2.4e-11 | --(1.3e-01) | 0.32(0.06) | 4.7e-07 | 0.18(0.07) | 6.0e-03 | 0.03(0.01) | 6.9e-05 | | | |
| 12 | rs110291686 | 19 | 43941099 | a/c | U2; ARL4D | 1.7e-15 | --(3.3e-01) | 0.20(0.04) | 1.4e-07 | 0.14(0.06) | 1.7e-02 | 0.04(0.01) | 1.3e-08 | | | |
| 13 | rs109082193 | 20 | 18413770 | a/g | ERCC8 | 1.3e-12 | ++++(7.0e-01) | -0.19(0.05) | 5.9e-04 | -0.11(0.05) | 3.9e-02 | -0.05(0.01) | 3.2e-09 | | | |
| 14 | rs385826566 | 20 | 39917111 | a/c | ADAMTS12 | 5.0e-22 | ??(1.0e+00) | -0.37(0.04) | 5.0e-22 | — | — | — | — | | | |
| 15 | rs133559545 | 21 | 8262352 | a/g | IGF1R | 2.4e-17 | ++?(1.0e+00) | -0.30(0.04) | 2.4e-17 | — | — | — | — | | | |
| 16 | rs41644660 | 26 | 19460476 | t/c | — | 1.1e-27 | ++++(3.2e-07) | -0.27(0.05) | 3.5e-08 | -0.76(0.07) | 9.3e-27 | — | — | | | |

¹Populations in order are MON, NOR, and HOL; "+" and "-" denote positive and negative substitution effects of the alternate allele. "?" indicates that the variant did not segregate in the respective population. The P-value of Cochran's Q-test for heterogeneity of the effect sizes across breeds is given in parentheses, and is significant if P < 0.05.



0.16 Mb upstream of *MEF2C* gene, and 0.3 Mb downstream of *TMEM161B* gene), and rs43430263 (BTA12:53241975, $P = 3.3 \times 10^{-14}$, *SLAIN1* gene).

Meta-Analysis Heterogeneity Among Three Breeds

As an example of the heterogeneity observed from a meta-analysis of the three breeds, we observed peaks from a combined Manhattan plot for fat and protein percentage (Figures 3A,B) and standardized the allelic substitution effects by the phenotypic standard deviation of fat and protein percentage meta-analysis results. We observed heterogeneity (Cochran's Q , $P < 0.05$) for 36 QTL from the meta-analysis in three breeds (Figure 3D). We then overlay observed breed specific QTL for these two traits and Holstein had 26 QTL in 24 chromosomes, Montbeliarde had 17 QTL in 14 chromosomes, and Normande had 16 QTL in 7 chromosomes (Figure 3C).

DISCUSSION

Our meta-analysis of association summary statistics for seven traits across three French dairy cattle breeds discovered 120 QTL including 13 QTL that had not been detected at $P < 1.03 \times 10^{-6}$ in the within-breed analyses. In agreement with previous studies (Pausch et al., 2017) our results show that combining GWAS summary data from several breeds increases the power of association studies. However, for lowly heritable traits, and for distinct populations, a meta-analysis may not be as robust

as the within-breed GWAS (Raven et al., 2014). For instance, within-breed GWAS showed associated SNP for fertility had no overlap between breeds probably due to small effects per locus (i.e., low power) and possibly the action of many, reasonably rare recessive lethal alleles. Holstein, with larger sample size, had associated SNP in strong LD and thus had stronger fertility predicted effects. In some instances, the meta-analysis increased the power, and consequently, the confidence interval for some fertility related QTL became visible, e.g., QTL on BTA2, 8 and 20 (Table 4). However, in other instances, the QTL detection power reduced when the QTL are private, e.g., the meta-analysis for fertility lost power compared to the within-breed GWAS when the Normande breed (with the smallest number of animals) was included in the analysis, e.g., QTL on BTA15 and 24 (Table 4).

One advantage of a meta-analysis over within-breed GWAS is that it is expected to result in more precision in QTL position due to the breakdown of long-range LD due to many generations of recombination since the breeds were separated (Raven et al., 2014; Pausch et al., 2016). In this study, we performed a within-breed GWAS first, followed by meta-analyses of GWAS summary statistics from three breeds using a weighted Z-score model to get the P -value for each SNP. In our study, the meta-analysis increased the P -value for some variants, probably because the QTL was not segregating across populations. Meta-analysis decreased P -value for some variants, and consequently, some QTL peaks became more distinct and narrower. For example, a region on BTA6:87296809 was strongly associated with protein yield and mapped near the protein-coding genes *ODAM* and

CSN3 in a 120 kb region. Bovine odontogenic, ameloblast-associated (*ODAM*) participates in structuring the extracellular matrix to attach epithelial cells to mineralized surfaces thus forming a protective seal that is antagonistic to bacterial invasion (Fouillen et al., 2017). On the other hand, Casein kappa (*CSN3*) is associated with daily milk yield (Bonfatti et al., 2011; Bartonova et al., 2012) as well as better cheese properties (Di Gregorio et al., 2017).

Strongly associated SNP ($P < 1 \times 10^{-50}$) were identified in all three French dairy breeds. Among these were several SNP that had previously been described such as those close to the *DGAT1-K232A* mutation on BTA14 (Grisart et al., 2002). These SNP were highly associated with all milk production traits in all three breeds. A very significant sequence variant at BTA5:93948357 (rs209372883) was associated with protein and fat in Holstein and Montbeliarde; it has previously been ascribed to a variant in *MGST1* (microsomal glutathione S-transferase 1; Raven et al., 2016), an upstream intron variant. *MGST1* is also an inflammation response gene which is highly expressed through pregnancy and lactation (Church et al., 2012). Previous studies in Japanese Black cattle (Wang et al., 2005) suggested upregulation of *MGST1* during adipocyte development in the longissimus muscle. Slightly upstream, a highly significant peak was centered within *SLC15A5*, a protein-coding gene. Our results suggest that both *MGST1* and *SLC15A5* may contain variants affecting milk production. Raven et al. (2014) identified a fat yield QTL within 3kbp of *MGST1* using a multi-breed analysis and we report a QTL which was within 2kbp of the genic region of *MGST1* using meta-analysis. Similar studies in Canadian Holstein, and using a similar density chip (50 k), reported a QTL 200 bp from that reported in our study (Li et al., 2010). The same Canadian Holstein study reported several fat yield candidate genes including *SLC2A3* and *GDF3* at 101.7–101.8 Mb and *LRP6*, *EMP1*, and *DUSP16* at 97–98 Mb on BTA5 (Li et al., 2010). Their results agrees with our meta-analysis results. Another very significant QTL for milk yield in Holstein was identified on BTA20 at ~38 Mb. The lead SNP of this QTL was located near the *PRLR* gene which has previously been associated with fertility in Holstein cows (Leyva-Corona et al., 2018; Marete et al., 2018b). The majority of the associated SNP with fertility, a low heritability trait, were 50 k chip variants. For instance, rs110992367, the lead SNP for a 3.3 Mb QTL for the success rate of insemination in Holstein cows was 0.38 Mb from the 5' flanking region of the *G2E3* gene. Meta-analysis heterogeneity for this QTL was observed in all three breeds ($P = 9.5 \times 10^{-3}$), but opposite in the effect direction between Holstein and the other two breeds. The heterogeneity is important because the *G2E3* gene is an essential protein-coding gene that prevents apoptotic death during embryonic development (Brooks et al., 2008). This gene has been linked to apoptosis and cellular stress in US beef cattle (Howard et al., 2014) and somatic cell score in Chinese Holstein (Wang et al., 2015). This gene is a good candidate affecting cow fertility.

All QTL identified by meta-analysis were not significant in all within-breed analysis (Tables 1–6, Figure 3C) probably due to low power to reach significance at $P < 1 \times 10^{-6}$ or lack of segregation in the breed (Sham and Purcell, 2014).

For instance, the variant mapping to the *PLAG1* gene only segregated in the Montbeliarde breed. However, most of the SNP had an effect in the same direction in all breeds for the more significant variants, e.g., BTA18:58067310 (*CEACAM18* gene, fat) and BTA10:51558941 (*FAM63B* gene, stature). This suggests that most detected QTL segregate in all three breeds even though they were not significant in the within-breed analysis. In this study, pre-selected sequence variants were directly genotyped or imputed from a large reference population genotyped with the EuroG10k chip, and thus were known with very high accuracy. This was evident from the high average allelic correlation ($r^2 = 0.97$) between real and imputed genotypes for the three breeds. In conclusion, many QTL segregated across the three French cattle breeds more so for highly heritable traits. The meta-analysis across the breeds using the within-breed association summary data increased the power of QTL detection. This QTL information can be incorporated in routine genomic evaluation in French populations. If an independent population of dairy cattle is genotyped with the EuroG10k chip, then the true causal variant underlying the QTL may be validated in independent population/breed. The next step of this study is to use independent Nordic cattle data to validate the QTL reported here.

AUTHOR CONTRIBUTIONS

AM analyzed data and wrote the manuscript. BG contributed to the design of the custom part of the EuroG10k chip, interpretation of results and revision of the manuscript. ML worked on the study design, secured project funding and revised the manuscript. SF contributed to the design of the custom part of the EuroG10k chip and the management of the marker information. GS interpreted the results and revised the manuscript. DB worked on the study design, interpretation of results and revision of the manuscript.

FUNDING

AM is a recipient of an Erasmus Mundus Ph.D. grant funded jointly by European Graduate School in Animal Breeding and Genetics (France), and Center for Genomic Selection in Animals and Plants (GenSAP) supported by Innovation Fund Denmark (grant 0603-00519B).

ACKNOWLEDGMENTS

We are grateful to French National Animal Breeding Database for providing phenotypic data, Valogene for access to genotypes and INRA for providing study environment. We also acknowledge Chris Hozé for initial genotype preparation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00522/full#supplementary-material>

REFERENCES

- Bartonova, P., Vrtkova, I., Kaplanova, K., and Urban, T. (2012). Association between CSN3 and BCO2 gene polymorphisms and milk performance traits in the Czech Fleckvieh cattle breed. *Genet. Mol. Res.* 11, 1058–1063. doi: 10.4238/2012.April.27.4
- Boichard, D., Boussaha, M., Capitan, A., Rocha, D., Hozé C., Sanchez, M. P., et al. (2018). “Experience from large scale use of the EuroGenomics custom SNP chip in cattle,” in *11th World Congress of Genetics Applied to Livestock Production* (Auckland), 675.
- Boichard, D., Chung, H., Dassonneville, R., David, X., Eggen, A. A., Fritz, S. S., et al. (2012a). Design of a bovine low-density snp array optimized for imputation. *PLoS ONE* 7:e34130. doi: 10.1371/journal.pone.0034130
- Boichard, D., Guillaume, F., Baur, A., Croiseau, P., Rossignol, M. N., Boscher, M. Y., et al. (2012b). Genomic selection in French dairy cattle. *Anim. Prod. Sci.* 52, 115–120. doi: 10.1071/AN11119
- Bonfatti, V., Cecchinato, A., Di Martino, G., De Marchi, M., Gallo, L., and Carnier, P. (2011). Effect of κ -casein B relative content in bulk milk on Montasio, Asiago, and Caciotta cheese yield using milk of similar protein composition. *J. Dairy Sci.* 94, 602–613. doi: 10.3168/jds.2010-3368
- Bouwman, A. C., Daetwyler, H. D., Chamberlain, A. J., Hurtado Ponce, C., Sargolzaei, M., Schenkel, F. S., et al. (2018). Meta-analysis of genome wide association studies for the stature of cattle reveals common genes that regulate size in mammals. *Nat. Gen.* 50, 362–367. doi: 10.1038/s41588-018-0056-5
- Brooks, W. S., Helton, E. S., Banerjee, S., Venable, M., Johnson, L., Schoeb, T. R., et al. (2008). G2E3 is a dual function ubiquitin ligase required for early embryonic development. *J. Biol. Chem.* 283, 22304–22315. doi: 10.1074/jbc.m803238200
- Chang, M., He, L., and Cai, L. (2018). An overview of genome wide association studies. *Comput. Syst. Biol.* 1754, 97–108. doi: 10.1007/978-1-4939-7717-8_6
- Church, P. C., Goscinski, A., and Lefèvre, C. (2012). EXP-PAC: Providing comparative analysis and storage of next generation gene expression data. *Genomics* 100, 8–13. doi: 10.1016/j.ygeno.2012.05.007
- Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics* 10, 101–129. doi: 10.2307/3001666
- Daetwyler, H. D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brøndum, R. F., et al. (2014). Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* 46, 858–865. doi: 10.1038/ng.3034
- Di Gregorio, P., Di Grigoli, A., Di Trana, A., Alabiso, M., Maniaci, G., Rando, A., et al. (2017). Effects of different genotypes at the CSN3 and LGB loci on milk and cheese-making characteristics of the bovine Cinisara breed. *Int. Dairy J.* 71:43105. doi: 10.1016/j.idairyj.2016.11.001
- Fouillen, A., Dos Santos Neves, J., Mary, C., Castonguay, J.-D., Moffatt, P., Baron, C., et al. (2017). Interactions of AMTN, ODAM and SCPPPQ1 proteins of a specialized basal lamina that attaches epithelial cells to tooth mineral. *Sci. Rep.* 24:46683. doi: 10.1038/srep46683
- Goddard, M. E., and Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10, 381–391. doi: 10.1038/nrg2575
- Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., et al. (2002). Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Res.* 12, 222–231. doi: 10.1101/gr.224202
- Hoff, J. L., Decker, J. E., Schnabel, R. D., and Taylor, J. F. (2017). Candidate lethal haplotypes and causal mutations in Angus cattle. *BMC Genomics.* 18:799. doi: 10.1186/s12864-017-4196-2
- Howard, J. T., Kachman, S. D., Snelling, W. M., Pollak, E. J., Ciobanu, D. C., Kuehn, L. A., et al. (2014). Beef cattle body temperature during climatic stress: a genome-wide association study. *Int. J. Biometeorol.* 58, 1665–1672. doi: 10.1007/s00484-013-0773-5
- Hu, Z.-L., Park, C. A., Reecy, J. M. (2016). Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Res.* 44, D827–D833. doi: 10.1093/nar/gkv1233
- Leyva-Corona, J. C., Reyna-Granados, J. R., Zamorano-Algandar, R., et al. (2018). Polymorphisms within the prolactin and growth hormone/insulin-like growth factor-1 functional pathways associated with fertility traits in Holstein cows raised in a hot-humid climate. *Trop. Anim. Health Prod.* doi: 10.1007/s11250-018-1645-0. [Epub ahead of print].
- Li, H., Wang, Z., Moore, S. S., Schenkel, F. S., and Stothard, P. (2010). “Genome-wide scan for positional and functional candidate genes affecting milk production traits in Canadian Holstein Cattle,” in *Proc 9th WCGALP, Genetics of Trait Complexes: Lactation - Lecture Sessions: 0535* (Leipzig). Available online at: <http://www.wcgalp.org>
- Marete, A., Lund, M. S., Boichard, D., and Ramayo-Caldas, Y. (2018a). A system-based analysis of the genetic determinism of udder conformation and health phenotypes across three French dairy cattle breeds. *PLoS ONE* 13:e0199931. doi: 10.1371/journal.pone.0199931
- Marete, A., Sahana, G., Fritz, S., Lefebvre, R., Barbat, A., Lund, M. S., et al. (2018b). Genome-wide association study for milking speed in French Holstein cows. *J. Dairy Sci.* 101, 6205–6219. doi: 10.3168/jds.2017-14067
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., et al. (2016). The Ensembl variant effect predictor. *Genome Biol.* 17:122. doi: 10.1186/s13059-016-0974-4
- Michot, P., Fritz, S., Barbat, A., Boussaha, M., Deloche, M. C., Grohs, C., et al. (2017). A missense mutation in PFAS (phosphoribosylformylglycinamide synthase) is likely causal for embryonic lethality associated with the MH1 haplotype in Montbéliarde dairy cattle. *J. Dairy Sci.* 100, 8176–8187. doi: 10.3168/jds.2017-12579
- Pausch, H., Emmerling, R., Grendler-Grandl, B., Fries, R., Daetwyler, D., and Goddard, M. E. (2017). Meta-analysis of sequence-based association studies across three cattle breeds reveals 25 QTL for fat and protein percentages in milk at nucleotide resolution. *BMC Genomics* 18:853. doi: 10.1186/s12864-017-4263-8
- Pausch, H., Emmerling, R., Schwarzenbacher, H., and Fries, R. (2016). A multi-trait meta-analysis with imputed sequence variants reveals twelve QTL for mammary gland morphology in Fleckvieh cattle. *Genet. Sel. Evol.* 48:14. doi: 10.1186/s12711-016-0190-4
- Raven, L. A., Cocks, B. G., and Hayes, B. J. (2014). Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. *BMC Genomics.* 15:62. doi: 10.1186/1471-2164-15-62
- Raven, L. A., Cocks, B. G., Kemper, K. E., Chamberlain, A. J., Jagt, C. J. V., Goddard, M. E., et al. (2016). Targeted imputation of sequence variants and gene expression profiling identifies twelve candidate genes associated with lactation volume, composition and calving interval in dairy cattle. *Mamm. Genome* 27, 81–97. doi: 10.1007/s00335-015-9613-8
- Sargolzaei, M., Chesnais, J. P., and Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15:478. doi: 10.1186/1471-2164-15-478
- Sham, P. C., and Purcell, S. M. (2014). Statistical power and significance testing in large-scale genetic studies. *Nat. Rev. Genet.* 15, 335–346. doi: 10.1038/nrg3706
- van den Berg, I., Boichard, D., and Lund, M. S. (2016). Comparing power and precision of within-breed and multibreed genome-wide association studies of production traits using whole-genome sequence data for 5 French and Danish dairy cattle breeds. *J. Dairy Sci.* 99, 8932–8945. doi: 10.3168/jds.2016-11073
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- VanRaden, P. M., and Wiggans, G. R. (1991). Derivation, calculation, and use of national animal model information. *J. Dairy Sci.* 74, 2737–2746. doi: 10.3168/jds.S0022-0302(91)78453-1
- Wang, X., Ma, P., Liu, J., Zhang, Q., Zhang, Y., Ding, X., et al. (2015). Genome-wide association study in Chinese Holstein cows reveal two candidate genes for somatic cell score as an indicator for mastitis susceptibility. *BMC Genet.* 16:111. doi: 10.1186/s12863-015-0263-3
- Wang, Y. H., Reverter, A., Mannen, H., Taniguchi, M., Harper, G. S., Oyama, K., et al. (2005). Transcriptional profiling of muscle tissue in growing Japanese

- Black cattle to identify genes involved with the development of intramuscular fat. *Aust. J. Exp. Agric.* 45, 809–820. doi: 10.1071/EA05058
- Willer, C. J., Li, Y., and Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190–2191. doi: 10.1093/bioinformatics/btq340
- Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011
- Zimin, A. V., Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., et al. (2009). A whole-genome assembly of the domestic cow. *Genome Biol.* 10:R42. doi: 10.1186/gb-2009-10-4-r42

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Marete, Guldbrandtsen, Lund, Fritz, Sahana and Boichard. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.