

# The impact of QTL sharing and properties on multi-breed GWAS in cattle: a simulation study

Irene van den Berg<sup>A,\*</sup>  and Iona M. MacLeod<sup>A,B</sup>

For full list of author affiliations and declarations see end of paper

**\*Correspondence to:**

Irene van den Berg  
Agriculture Victoria, AgriBio, Centre for  
AgriBioscience, 5 Ring Road, Bundoora,  
Vic. 3083, Australia  
Email:  
[irene.vandenberg@agriculture.vic.gov.au](mailto:irene.vandenberg@agriculture.vic.gov.au)

**Handling Editor:**

Sue Hatcher

**Received:** 14 December 2022

**Accepted:** 13 March 2023

**Published:** 6 April 2023

**Cite this:**

van den Berg I and MacLeod IM (2023)  
*Animal Production Science*, **63**(10–11),  
996–1007.  
doi:[10.1071/AN22460](https://doi.org/10.1071/AN22460)

© 2023 The Author(s) (or their  
employer(s)). Published by  
CSIRO Publishing.

This is an open access article distributed  
under the Creative Commons Attribution-  
NonCommercial-NoDerivatives 4.0  
International License (CC BY-NC-ND).

OPEN ACCESS

## ABSTRACT

**Context.** Genome-wide association studies (GWAS) and meta-analyses can be used to detect variants that affect quantitative traits. Multi-breed GWAS may lead to increased power and precision compared with within-breed GWAS. However, not all causal variants segregate in all breeds, and variants that segregate in multiple breeds may have different allele frequencies in different breeds. It is not known how differences in minor allele frequency (MAF) affect multi-breed GWAS and meta-analyses. **Aims.** Our aim was to study the impact of differences in MAF at causal variants on mapping power and precision. **Methods.** We used real imputed sequence data to simulate quantitative traits in three dairy cattle breeds. Causal variants (QTN) were simulated according to the following three scenarios: variants with a similar MAF in all breeds, variants with a lower MAF in one breed than the other, and variants that each only segregated in one of the breeds. We analysed the simulated quantitative traits with three methods to compare mapping power and precision: within-breed GWAS, multi-breed GWAS and meta-analysis. **Key results.** Our results indicated that the multi-breed analyses (multi-breed GWAS or meta-analysis) detected similar or more QTN than did within-breed GWAS, with improved mapping precision in most scenarios. However, when MAF differed between breeds, or variants were breed specific, the advantage of the multi-breed analyses over within breed GWAS decreased. Regardless of the type of QTN (similar MAF in all breeds, different MAF in different breeds, or only segregating in one breed), multi-breed GWAS and meta-analyses performed similar or better than did within-breed GWAS, demonstrating the benefits of multi-breed GWAS. We did not find large differences between the results obtained with the meta-analysis and multi-breed GWAS, confirming that a meta-analysis can be a suitable approximation of a multi-breed GWAS. **Conclusions.** Our results showed that multi-breed GWAS and meta-analysis generally detect more QTN with improved precision than does within-breed GWAS, and that even with differences in MAF, multi-breed analyses did not perform worse than within-breed GWAS. **Implications.** Our study confirmed the benefits of multi-breed GWAS and meta-analysis.

**Keywords:** allele frequency, dairy cattle, GWAS, meta-analysis, multi-breed, QTL detection, quantitative traits, within breed.

## Introduction

Genome-wide association studies (GWAS) and meta-analyses can be used to detect variants that affect quantitative traits (Bouwman *et al.* 2018; Jiang *et al.* 2019) and to select predictive markers that can improve the accuracy of genomic prediction (Brøndum *et al.* 2015; VanRaden *et al.* 2017; Xiang *et al.* 2019). Multi-breed GWAS with sequence data may lead to increased power and precision compared with within-breed GWAS (van den Berg *et al.* 2016a). The inclusion of predictive sequence markers selected from GWAS can be especially beneficial for across-breed (Raymond *et al.* 2018a) and multi-breed genomic prediction (Raymond *et al.* 2018b). Furthermore, variants selected from a multi-breed GWAS can result in higher prediction accuracies than variants selected from within-breed GWAS (van den Berg *et al.* 2016b). However, not all causal variants

segregate in all breeds (Raven *et al.* 2014; Kemper *et al.* 2015), and variants that segregate in multiple breeds may not have the same allele frequency in each breed. Furthermore, some regions may contain multiple causal variants that do not always segregate in all breeds. It is not known how differences in minor allele frequency (MAF) affect multi-breed GWAS and meta-analyses. Therefore, we simulated scenarios where causal variants had similar or different MAF in three dairy cattle breeds, and compared detection power and precision of within-breed GWAS, multi-breed GWAS and multi-breed meta-analysis.

## Materials and methods

### Genotypes

We used real genotype data for 66 739 Holstein (HOL), 13 398 Jersey (JER) and 5536 Australian Red (RED) individuals, including imputed sequence data and Illumina BovineHD BeadChip genotypes (HD). A detailed description of the imputation pipeline has been provided in van den Berg *et al.* (2022). Animals were genotyped at various single-nucleotide polymorphism (SNP) array densities. First, animals imputed with a low-density SNP panel were imputed to the Illumina Bovine 50K panel by using a mixed-breed imputation reference set, containing 14 722 HOL, JER and RED animals. Subsequently, all animals were imputed from 50K to HD, with a reference population of 2700 animals (HOL, JER and RED), and HD to whole genome-sequence level, by using a reference population of 5490 *Bos taurus* cattle from Run8 of the 1000 Bulls Genome Project (Daetwyler *et al.* 2014; Hayes and Daetwyler 2019). Imputation up to HD was performed using Fimpute v.3 (Sargolzaei *et al.* 2014) and sequence imputation was performed using Minimac4 (Das *et al.* 2016) and Eagle v.2.4.1 (Loh *et al.* 2016). All imputation was undertaken using multi-breed reference populations, because previous studies (Bouwman and Veerkamp 2014; Brøndum *et al.* 2014; Pausch *et al.* 2017) have shown that multi-breed imputation tends to result in a very similar or slightly higher imputation accuracy than does within-breed imputation. For this study, we used sequence variants only on Chromosome 1 and the 616 807 HD variants on Chromosomes 1–29. After removing sequence variants with an imputation  $r^2$  computed by Minimac4 lower than 0.4, there were 1 277 974 sequence variants retained on Chromosome 1, with a minor allele frequency (MAF) larger than 0 (for all breeds considered together). We used GCTA (Yang *et al.* 2011) to make a genomic relationship matrix (GRM) of the HD genotypes and performed a principal-component analysis (PCA) to check the breed identity present in the dataset. Based on the PCA (Supplementary material Fig. S1), we removed HOL with  $PC1 > 0.0001$  or  $PC2 > 0.005$ , JER with  $PC1 < 0.0055$  or  $PC2 > 0.005$  and RED with

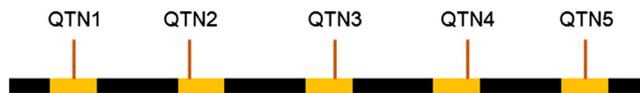
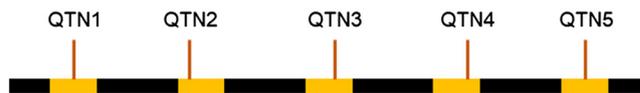
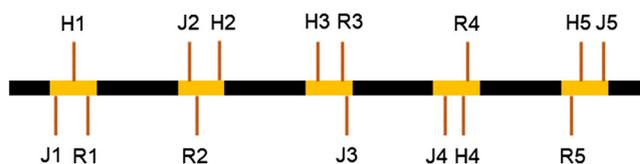
$PC1 > 0.004$  or  $PC2 < 0.005$ . After this, there were 66 710 HOL, 13 291 JER and 5385 RED remaining.

### Simulation

To simulate the causal variants, or quantitative-trait nucleotides (QTN), we first divided Chromosome 1 in five causal regions of 11.6 Mb each, each separated by 20 Mb windows. The causal regions were located between 10 565 483 and 22 158 125 bp, 42 158 125 and 53 750 767 bp, 73 750 767 and 85 343 409 bp, 105 343 409 and 116 936 051 bp and 136 936 051 and 148 528 693 bp. We simulated 100 QTN by randomly selecting HD variants from Chromosomes 2–29, and 5 or 15 QTN on Chromosome 1 according to the following scenarios (Fig. 1):

- Similar MAF: 5 QTN on Chromosome 1, one per causal region, with similar MAF in all breeds:
  - Common:  $MAF > 0.01$  in all breeds
  - Low:  $MAF 0.001–0.01$  in all breeds
  - Rare:  $MAF > 0–0.001$  in all breeds
- Different MAF: 5 QTN on Chromosome 1, one per causal region, QTN were either rare ( $MAF > 0–0.001$ ) or fixed ( $MAF = 0$ ) in one breed and had a  $MAF > 0.001$  in the other two breeds:
  - RareH: rare ( $MAF > 0–0.001$ ) in HOL, and  $MAF > 0.001$  in JER and RED
  - RareJ: rare ( $MAF > 0–0.001$ ) in JER, and  $MAF > 0.001$  in HOL and RED
  - RareR: rare ( $MAF > 0–0.001$ ) in RED, and  $MAF > 0.001$  in HOL and JER
  - FixedH: fixed ( $MAF = 0$ ) in HOL, and  $MAF > 0.001$  in JER and RED
  - FixedJ: fixed ( $MAF = 0$ ) in JER, and  $MAF > 0.001$  in HOL and RED
- Breed-specific: 15 QTN on Chromosome 1, three per causal region, one segregating only in HOL ( $MAF > 0.001$  in HOL, and  $MAF = 0$  in JER and RED), one only in JER ( $MAF > 0.001$  in JER, and  $MAF = 0$  in HOL and JER) and one only in RED ( $MAF > 0.001$  in RED and  $MAF = 0$  in HOL and JER).

Phenotypes based on the selected QTN were simulated using the ‘simu-qt’ and ‘simu-causal-loci’ functions in GCTA (Yang *et al.* 2011), with a heritability of 0.1, 0.3 and 0.5. Each scenario was repeated 10 times. In all scenarios, the simulated effect was the same in all breeds. In the different MAF scenarios, we did not simulate a scenario where QTN were fixed in RED and had a  $MAF > 0.001$  in HOL and JER, because insufficient variants fulfilled this MAF requirement. We randomly selected 5385 HOL and 5385 JER (same sample size as REDs) in each scenario to equalise detection power across breeds, so that the GWAS results were not confounded by very different breed proportions.

**(1) Similar MAF****(2) Different MAF****(3) Breed specific**

Scenario	Holstein	Jersey	Australian Red
Common	>0.01	>0.01	>0.01
Low	0.001–0.01	0.001–0.01	0.001–0.01
Rare	>0–0.001	>0–0.001	>0–0.001

Scenario	Holstein	Jersey	Australian Red
RareH	>0–0.001	>0.001	>0.001
RareJ	>0.001	>0–0.001	>0.001
RareR	>0.001	>0.001	>0–0.001
FixedH	0	>0.001	>0.001
FixedJ	>0.001	0	>0.001

Scenario	Holstein	Jersey	Australian Red
OnlyH	>0.001	0	0
OnlyJ	0	>0.001	0
OnlyR	0	0	>0.001

**Fig. 1.** Overview of simulated scenarios. Causal regions are indicated in yellow. Scenario 1, similar minor allele frequency (MAF): five quantitative trait nucleotides (QTN) on Chromosome 1 (QTN1–5), one per causal region, with similar MAF in all breeds (common (>0.01), low (0.001–0.01) or rare (>0–0.001)); Scenario 2, different MAF: five QTN on Chromosome 1 (QTN1–5), one per causal region, QTN were rare (MAF > 0–0.001) in or fixed (MAF = 0) in one breed and had a MAF > 0.001 in the other breeds; Scenario 3, breed-specific: 15 QTN on Chromosome 1, three per causal region, one segregating only in Holstein (H1–5), one in Jersey (J1–5), and one in Australian Red (R1–5).

## Association mapping methods

For each repeat, we performed a GWAS within each breed, a multi-breed GWAS combining the three breeds and a meta-analysis using the METAL software (Willer *et al.* 2010) that combined the three within breed GWAS (METAL). GCTA (Yang *et al.* 2011) was used for the within breed and multi-breed GWAS, fitting a GRM in all GWAS, and a breed effect in the multi-breed GWAS. For each GWAS, a GRM was constructed following Yang *et al.* (2011) on the basis of all autosomal HD variants and all individuals included in that GWAS (i.e. the within-breed GWAS used a GRM constructed using only genotypes of individuals of that breed, whereas the multi-breed GWAS used a multi-breed GRM). For the meta-analysis, we used the weighted Z-score model in METAL (Willer *et al.* 2010), using the  $P$ -value, direction of effect and sample size from the within-breed GWAS as input parameters. Subsequently, we did a conditional and joint analysis (COJO; Yang *et al.* 2012) on the summary statistics of the within-breed and multi-breed GWAS and the meta-analysis, to estimate the number of independent QTN signals.

## Mapping-method evaluation

For the evaluation of the mapping methods, we considered the sequence variants only on Chromosome 1. All variants with a

$P$ -value  $\leq 5 \times 10^{-8}$  were declared significant. Quantitative trait loci (QTL) intervals were constructed by first ranking all variants from smallest and largest  $P$ -value, and then grouping variants within 0.5 Mb distance of each other and a  $-\log_{10}(P)$  value of at least 2/3rd that of the most significant variant in the interval. The interval size is then defined by the minimum and maximum position of all variants included in the QTL interval.

To assess the three mapping methods (within-breed GWAS, multi-breed GWAS and meta-analysis), we evaluated the results on the basis of the following criteria:

- nQTN\_SIGN = the number of significant QTN
- nQTN\_QTL = the number of QTN located in QTL intervals
- nQTN\_COJO = number of QTN selected by COJO
- size\_QTL = the size of QTL intervals
- distQTN\_TOP = the distance between each QTN located in a QTL interval and the most significant variant in that QTL interval
- distQTN\_COJO = the distance between each QTN and the closest COJO variant (only if a COJO variant was selected in the causal region in which the QTN was located)
- propQTL\_false = the proportion of QTL intervals that did not contain a QTN
- nCOJO\_causal = the number of variants selected by COJO per causal region

- nCOJO\_window = the number of variants selected by COJO per window between causal regions

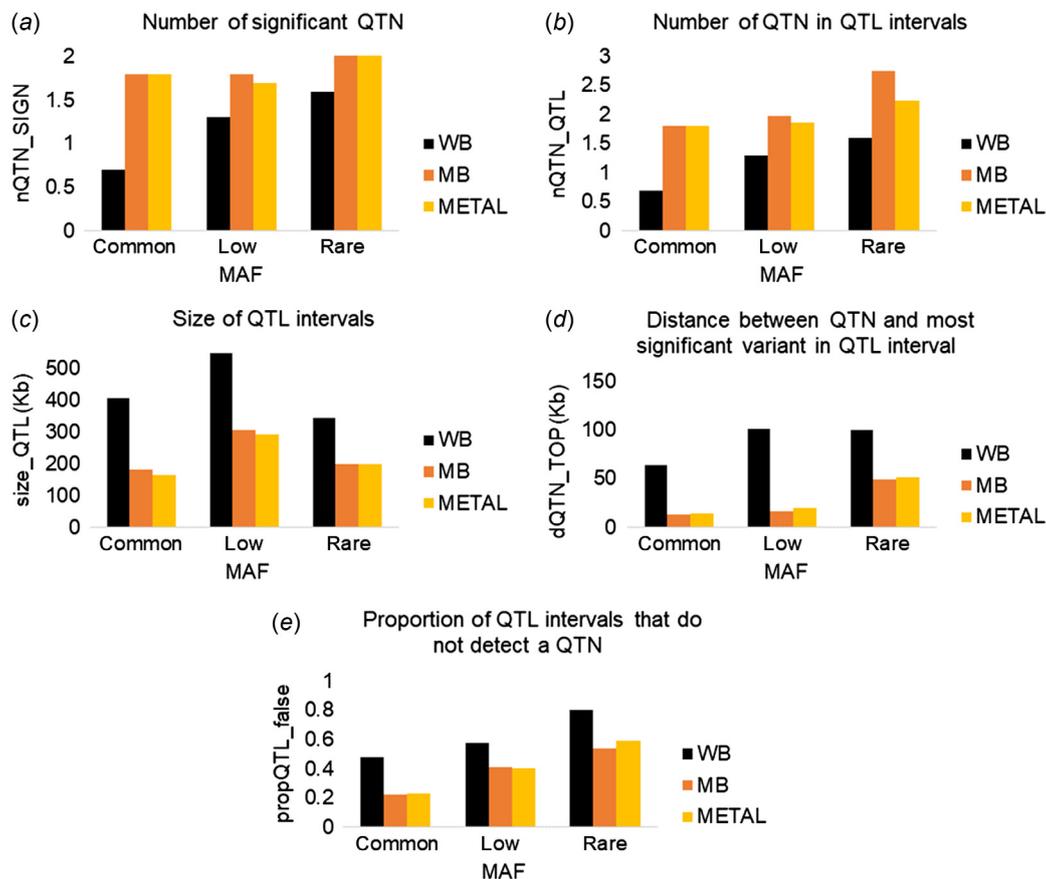
For the within-breed GWAS, nQTN\_SIGN, nQTN\_QTL and nQTN\_COJO were calculated as the number of unique QTN detected in the three within-breed GWAS (i.e. if the same QTN was significant both in HOL and in JER, it counted only as 1, but if one QTN was significant in HOL and a different QTN in JER, it counted as 2). Similarly, nCOJO\_causal and nCOJO\_window represent the number of unique variants detected by COJO per causal region (or window); hence, if a COJO selects a different variant in HOL than in JER in the same region (window), nCOJO\_causal (nCOJO\_window) equals 2. When comparing between two analyses (i.e. within-breed GWAS and multi-breed GWAS), distQTN\_TOP was calculated only if there was a QTL interval containing the QTN in both analyses. Similarly, distQTN\_COJO was estimated only when, in both analyses, COJO selected a variant in the window in which the QTN was simulated.

## Results

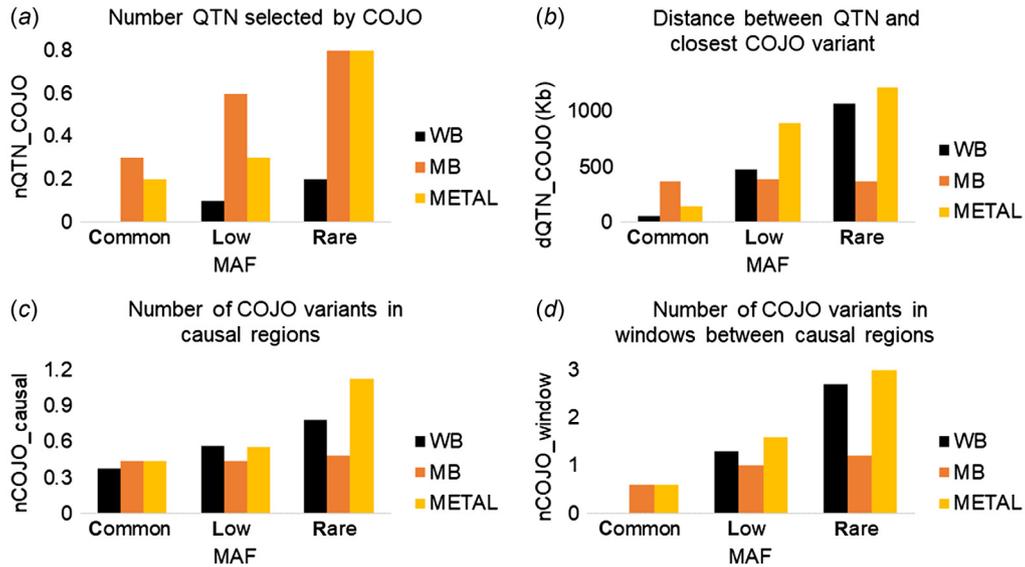
In most scenarios, the multi-breed GWAS and meta-analysis both detected similar or more QTN than did within-breed GWAS, with more precise mapping. The overall trend was the same regardless of heritability. Therefore, here we show more details for the results obtained with a heritability of 0.3, and full results, including the scenarios with heritabilities of 0.1 and 0.5, can be found in Supplementary material Tables S1–S3.

### Similar MAF

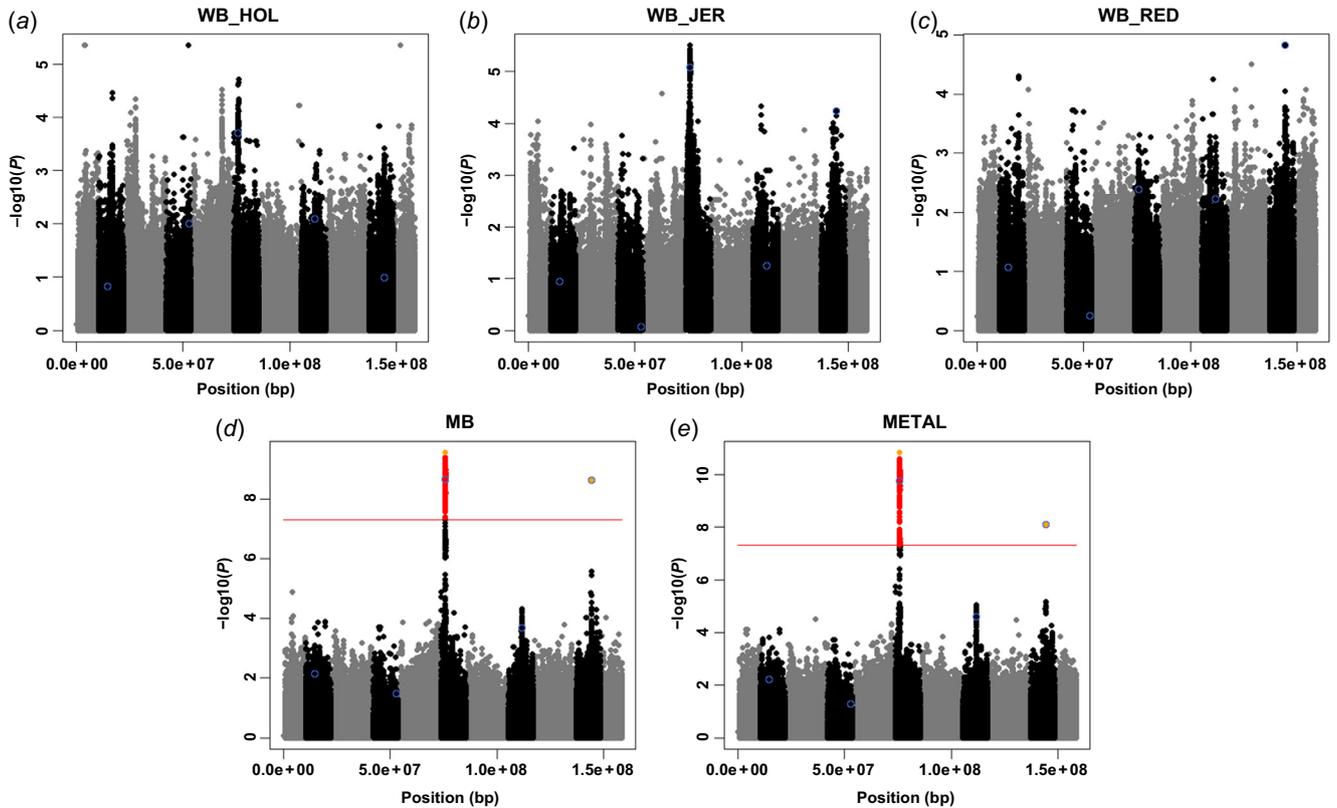
When the simulated QTN had a similar MAF in the three breeds, the multi-breed GWAS and meta-analysis both resulted in improved power (higher nQTN\_SIGN, nQTN\_QTL) and precision (smaller size\_QTL and distQTN\_TOP) compared with within-breed GWAS (Fig. 2). For example, with a heritability of 0.3 and a low MAF, the within-breed GWAS detected 1.3 QTN in the QTL intervals, with an average



**Fig. 2.** Summary of GWAS results for simulated QTN that had either common, low or rare minor allele frequencies in all breeds. Plots show (a) number of significant QTN, (b) number of significant QTN in QTL intervals, (c) size of QTL intervals, (d) distance between QTN and most significant variant in QTL interval and (e) proportion of QTL intervals that do not detect a QTN. The trait heritability was 0.3. There were three different analyses: within-breed GWAS (WB), multi-breed GWAS (MB) and meta-analysis (METAL).



**Fig. 3.** Summary of GWAS results for simulated QTN with common, low or rare minor allele frequencies in all breeds. Plots show (a) number of QTN selected by COJO, (b) distance between QTN and closest COJO variant, (c) number of COJO variants in causal regions, and (d) number of COJO variants in windows between causal regions. The trait heritability was 0.3.



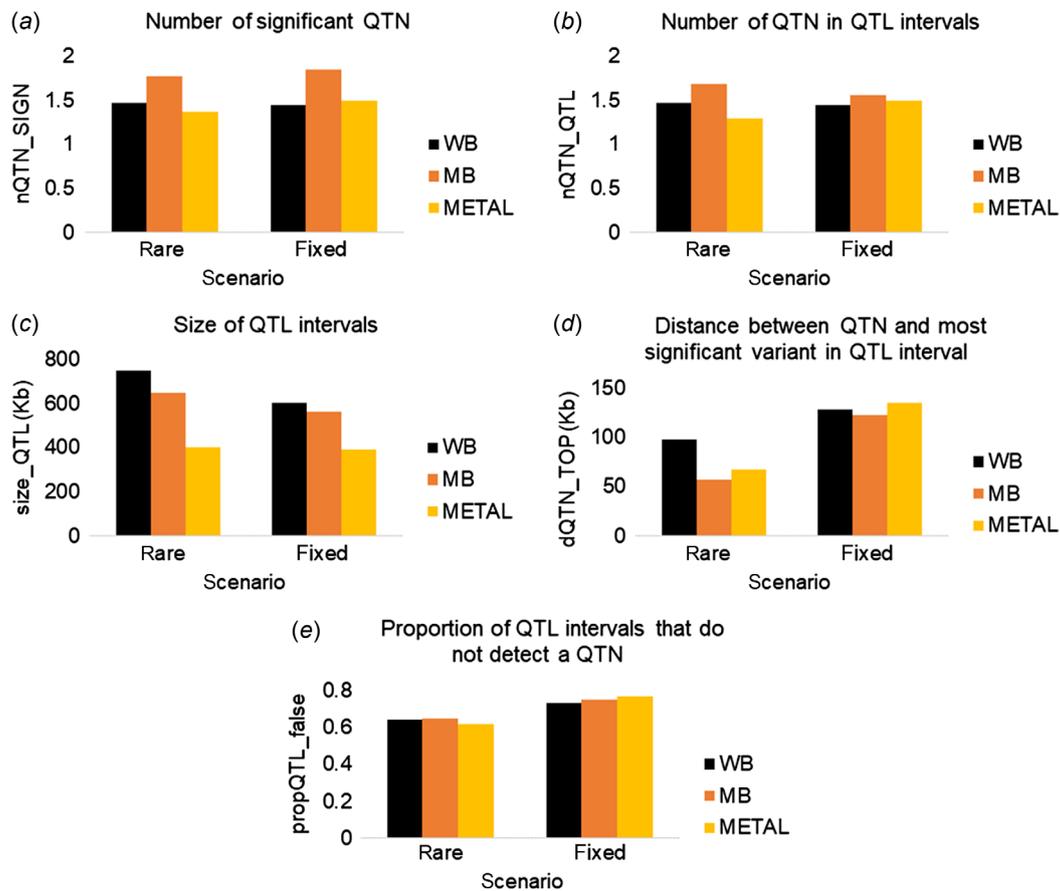
**Fig. 4.** Manhattan plots of GWAS results for simulated QTN with common, low or rare minor allele frequencies in all breeds. Plots show (a) within-breed Holstein (WB\_HOL), (b) within-breed Jersey (WB\_JER), (c) within-breed Australian Red (WB\_RED), (d) multi-breed (MB), and (e) meta-analysis (METAL). The trait heritability was 0.3. The red line indicates a significance threshold of  $P = 5 \times 10^{-8}$ , blue circles indicate simulated causal variants, red dots indicate variants in QTL regions, and orange dots variants selected by COJO.

interval size of 548 Kb, while the multi-breed GWAS and meta-analysis detected 1.8 QTN and 1.7 QTN, with an average interval size of 306 and 292 Kb respectively. For scenarios with a heritability of 0.1 and 0.3, the proportion of false positives ( $\text{propQTL\_false}$ ) was lower in the multi-breed GWAS and meta-analysis than within breed, while there was no consistent difference with a heritability of 0.5 (Table S1, Fig. 2). While  $\text{nQTN\_COJO}$  was larger in the multi-breed GWAS and meta-analysis than within breed, there was no consistent difference for  $\text{distQTN\_COJO}$ ,  $\text{nCOJO\_causal}$  and  $\text{nCOJO\_window}$  (Table S1, Fig. 3). Fig. 4 shows details of a repeat of the scenario where all QTN had a common MAF, with a heritability of 0.3, where none of the within breed GWAS detected any significant variants, while the multi-breed GWAS and meta-analysis detected 2 QTN in QTL intervals, and 1 QTN was selected by COJO.

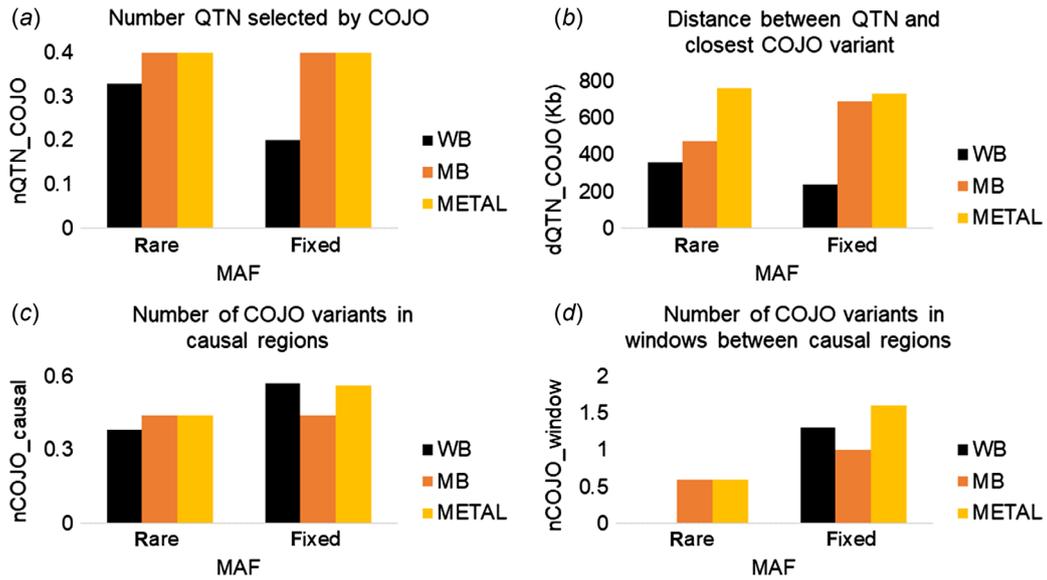
### Different MAF

In the different MAF scenarios, the multi-breed GWAS resulted in improved power (higher  $\text{nQTN\_SIGN}$ ,  $\text{nQTN\_QTL}$

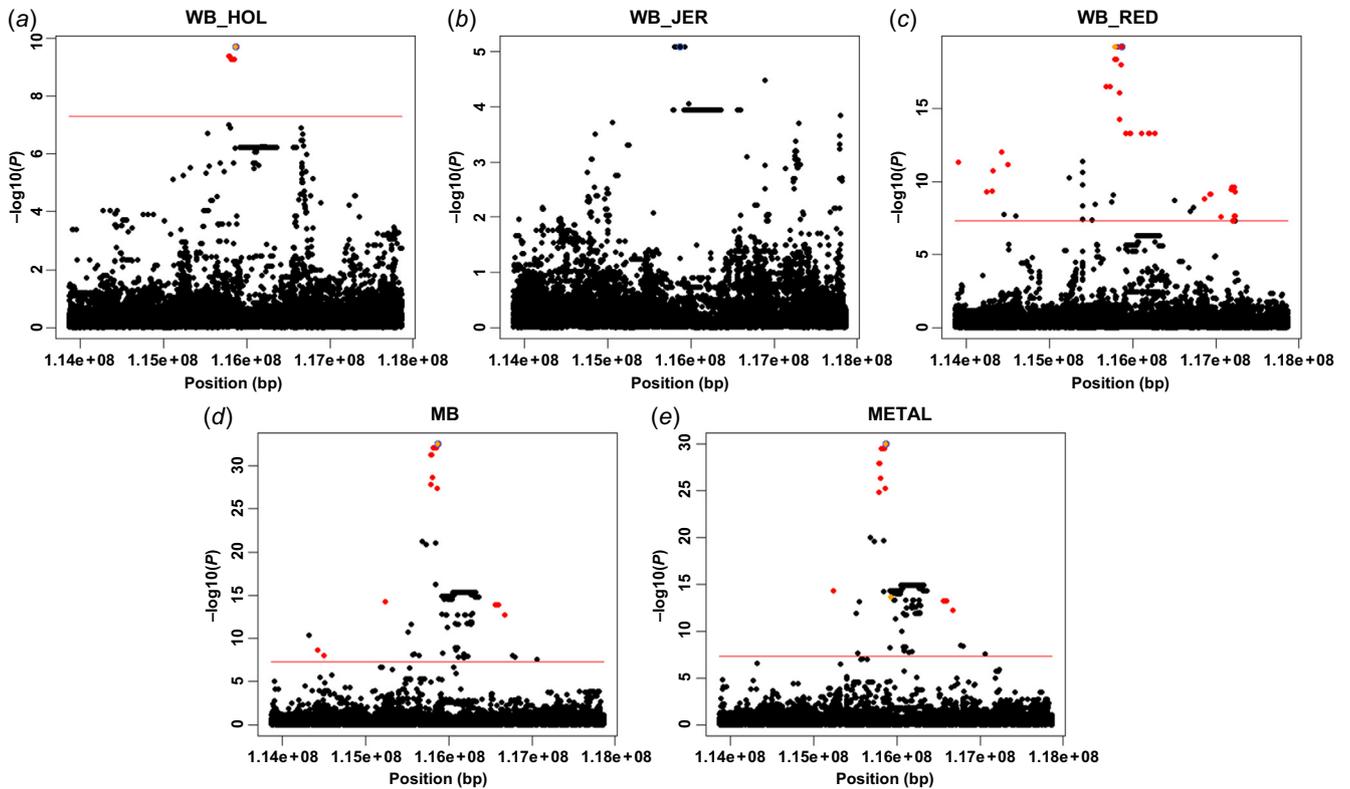
and  $\text{nQTN\_COJO}$ ), while there was no consistent difference between the within-breed GWAS and meta-analysis (Figs 5, 6, Table S2). For example, with a heritability of 0.3 and QTN with a rare MAF in one breed,  $\text{nQTN\_SIGN}$  equalled 1.77 in the multi-breed GWAS, 1.47 in the within breed GWAS and 1.37 in the meta-analysis. In the multi-breed and meta-analysis precision improved (smaller  $\text{size\_QTL}$  and  $\text{distQTN\_TOP}$ ) compared with within breed (Fig. 5), whereas the additional COJO analyses resulted in no consistent difference in precision observed for  $\text{distQTN\_COJO}$ ,  $\text{nCOJO\_causal}$  and  $\text{nCOJO\_window}$  (Fig. 6, Table S2). While differences were small,  $\text{propQTL\_false}$  was slightly smaller in the meta-analysis than in either of the GWAS (Fig. 5). Fig. 7 shows an example of one replicate of the rareJ scenario, highlighting QTN with a MAF of 0.004, 0.0008 and 0.008 in HOL, JER and RED respectively. This QTN was detected in the within-breed GWAS for both HOL ( $P = 2.0 \times 10^{-10}$ ) and RED ( $P = 5.9 \times 10^{-20}$ ), but was not significant in JER ( $P = 8.3 \times 10^{-6}$ ). The multi-breed GWAS or meta-analysis resulted in decreased  $P$ -values for the causal variant, down



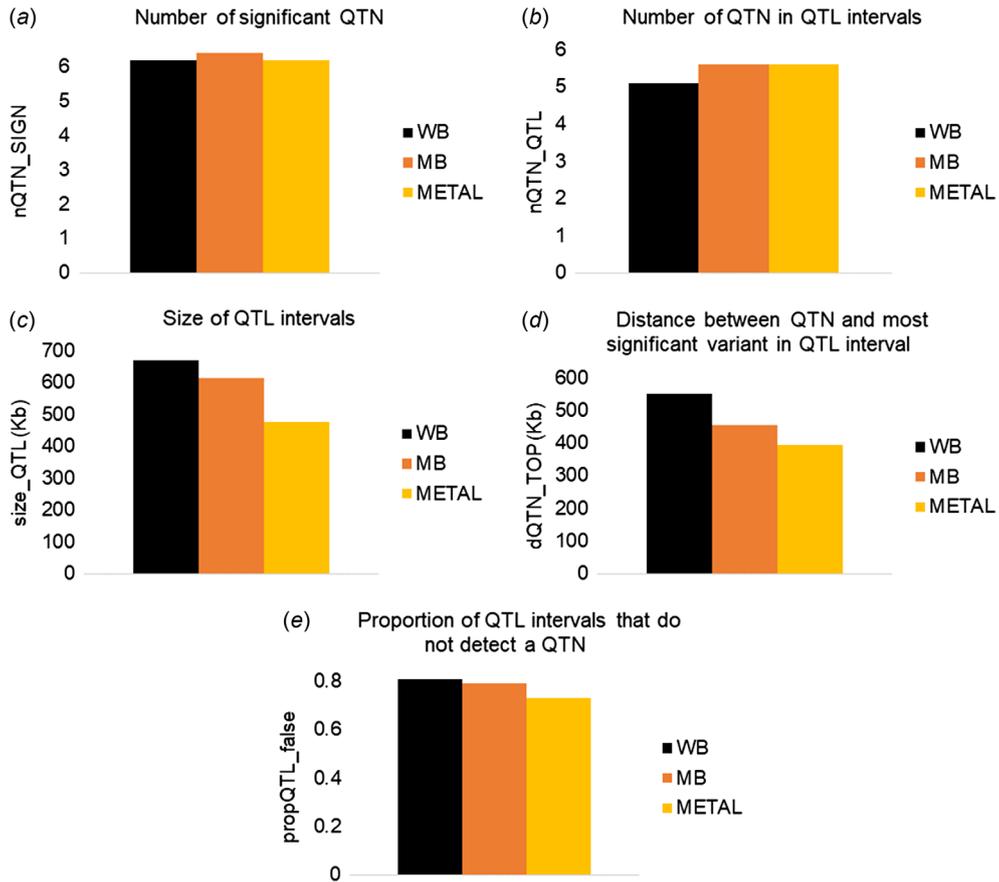
**Fig. 5.** Summary of GWAS results for simulated QTN that were rare or fixed in one breed but had a larger minor allele frequency in the other two breeds. Plots show (a) number of significant QTN, (b) number of QTN in QTL intervals, (c) size of QTL intervals, (d) distance between QTN and most significant variant in QTL interval, and (e) proportion of QTL intervals that do not detect a QTN. The trait heritability was 0.3 and three different analyses were implemented: within-breed GWAS (WB), multi-breed GWAS (MB) and meta-analysis (METAL).



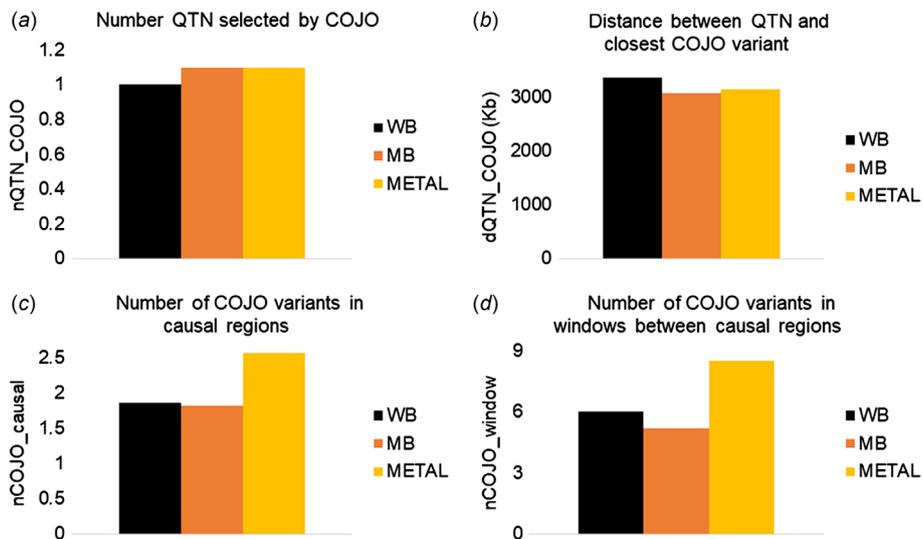
**Fig. 6.** Summary of GWAS results for simulated QTN that were rare or fixed in one breed and had a larger minor allele frequency in the other two breeds. Plots show (a) number of QTN selected by COJO, (b) distance between QTN and closest COJO variant, (c) number of COJO variants in causal regions, and (d) number of COJO variants in windows between causal regions. The trait heritability was 0.3.



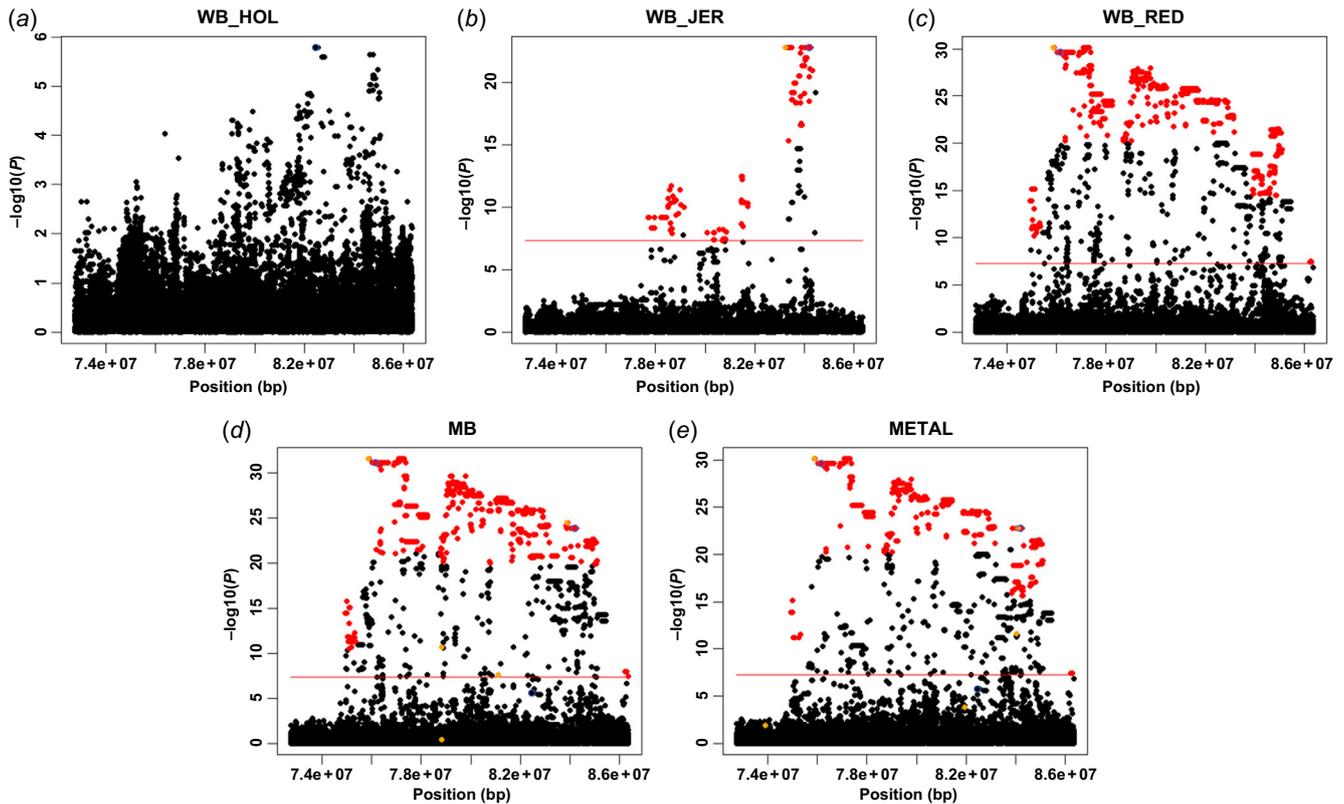
**Fig. 7.** Manhattan plots of GWAS results for a simulated QTN that had a lower minor allele frequency in Jersey than in Holstein and Red. The plots include (a) within-breed Holstein (WB\_HOL), (b) within-breed Jersey (WB\_JER), (c) within-breed Australian Red (WB\_RED), (d) multi-breed GWAS (MB) and (e) meta-analysis (METAL). The trait heritability was 0.3. The red line indicates a significance threshold of  $P = 5 \times 10^{-8}$ , blue circles indicate simulated causal variant, red dots indicate variants in QTL regions, and orange dots variants selected by COJO.



**Fig. 8.** Summary of GWAS results for breed-specific QTN, showing (a) number of significant QTN, (b) number of QTN in QTL intervals, (c) size of QTL intervals, (d) and distance between QTN and most significant variant in QTL interval, and (e) proportion of QTL intervals that do not detect a QTN. The trait heritability was 0.3.



**Fig. 9.** Summary of GWAS results for breed-specific QTN, showing (a) number of QTN selected by COJO, (b) distance between QTN and closest COJO variant, (c) number of COJO variants in causal regions, and (d) number of COJO variants in windows between causal regions. The trait heritability was 0.3 and analyses included within-breed GWAS (WB), multi-breed GWAS (MB) and meta-analysis (METAL).



**Fig. 10.** Manhattan plots of GWAS results for three breed-specific QTN. The plots include (a) within-breed Holstein (WB\_HOL), (b) within breed Jersey (WB\_JER), (c) within breed Australian Red (WB\_RED), (d) multi-breed GWAS (MB) and (e) meta-analysis (METAL). The trait heritability was 0.3. The red line indicates a significance threshold of  $P = 5 \times 10^{-8}$ , blue circles indicate simulated causal variant, red dots indicate variants in QTL regions, and orange dots variants selected by COJO.

to  $3.2 \times 10^{-33}$  and  $9.7 \times 10^{-31}$  respectively. Furthermore, while the QTN was the most significant variant in all of the analyses, there were nine variants with the same  $P$ -value in the within-breed JER and RED GWAS, and only three in the HOL GWAS and multi-breed and meta-analyses.

### Breed specific

When causal variants were specific to one breed only, there were only minimal, inconsistent differences in  $nQTN\_SIGN$ ,  $nQTN\_QTL$ ,  $nQTN\_COJO$ ,  $distQTN\_COJO$  and between within-breed and multi-breed analyses (Figs 8, 9, Table S3). Smaller values of  $size\_QTL$ ,  $distQTN\_TOP$  and  $propQTL\_false$  were obtained with the multi-breed GWAS and meta-analysis than within breed (Figs 8, 9), and  $nCOJO\_causal$  and  $nCOJO\_windows$  were larger in the meta-analysis than in either the within- or multi-breed GWAS (Fig. 9). Fig. 10 shows a replicate for one QTL region where the two different QTN segregating in JER and RED were detected both by the within-breed GWAS, multi-breed GWAS and meta-analysis. Visually, the JER and RED QTN appeared to be part of the same QTN in the multi-breed GWAS and meta-analysis.

The third simulated QTN segregating in HOL was not detected in any of the analyses.

### Discussion

Our results showed that multi-breed GWAS and meta-analysis tend to detect more QTN with improved precision than does within-breed GWAS, if QTN have a similar MAF. This is in agreement with previous studies in dairy cattle with real production phenotypes (van den Berg *et al.* 2016a, 2020; Marete *et al.* 2018; Teissier *et al.* 2018), which suggested that multi-breed analyses had a higher power and precision to detect some QTL than did within-breed analyses. However, in real data, the causal variants are seldom known a priori, so we cannot know what proportion of causal variants are shared or what their MAF is in different breeds. In different MAF and breed-specific scenarios, the advantage of the multi-breed analyses over within-breed GWAS decreased. When QTN segregate only in one breed, a multi-breed GWAS or meta-analysis did not lead to an increase in power for that QTN, explaining the limited advantage of multi-breed analyses

over within-breed analyses in the breed-specific scenarios. In theory, precision might still be improved because LD is conserved over shorter distances across breeds than within breed (De Roos *et al.* 2008). This may occur when the QTN is not segregating in one breed, but variants nearby that are in high LD with the QTN in another breed in which it segregates, do segregate in the first breed. In this scenario, the variants near the QTN may be highly significant in the within-breed GWAS of the breed in which the QTN segregates, but not in the other breed. Consequently, the *P*-values of those variants close to the QTN would become less significant in the multi-breed GWAS, resulting in a narrower peak. Indeed, the size of QTL intervals and distance between each QTN located in a QTL interval and the most significant variant in that QTL interval were smaller for the multi-breed GWAS and meta-analysis than for the within-breed GWAS, even in the breed-specific scenarios. Because the majority of causal variants (QTN) in dairy cattle are unknown, we cannot verify how realistic the simulated scenarios are. However, there are examples of shared QTN across breeds (Gautier *et al.* 2007), as well as QTN in beef cattle that are in and around the gene encoding myostatin but are completely different mutations (Bellinche *et al.* 2005). On the basis of the allele frequencies of the sequence variants observed in the breeds in our simulation (Fig. S2), there are substantially more variants (44%) that had a similar MAF in all three breeds, than variants that were rare (15%) or fixed (11%) in one breed but not in the other two breeds, or were breedspecific (9%). Hence, it seems realistic that some QTN are likely to fall into each of these categories. Regardless of the type of QTN, multi-breed GWAS and meta-analyses performed similar or better than within-breed GWAS, demonstrating the benefits of multi-breed GWAS.

A previous study using real data showed that unbalanced sample size between different breeds contributing to a multi-breed GWAS appeared to result in a GWAS being dominated by the breed with largest sample size (van den Berg *et al.* 2016a), with QTL detected within breeds with smaller population sizes overshadowed in the multi-breed GWAS by a nearby QTL detected in the breed with largest population size. In our study, we wanted to investigate the effect of MAF independent of differences in sample sizes. In reality, it will be generally preferable to maximise mapping power and use the maximum dataset. The weights used in a meta-analyses could be used to account for differences in sample size. A further study could investigate how differences in MAF at the QTN in combination with differences in sample size affect a multi-breed GWAS and meta-analysis.

Our simulation assumed that QTN have the same effect in all breeds. However, given the small number of confirmed causal mutations, we do not know whether QTN generally have the same effect in different breeds or not. QTN involved in gene  $\times$  gene or gene  $\times$  environment interactions, may have different effects in different breeds, which would reduce the power of multi-breed GWAS. Results from a multi-breed

GWAS of fat percentage and protein percentage in dairy cattle showed that the majority of QTL had the same direction of effect in all within-population GWAS they were detected (van den Berg *et al.* 2020). This suggests that, while the magnitude of effects may differ among breeds, QTN are likely to at least have the same direction of effect in multiple populations of the same species.

A caveat of this study is that we included all QTN in the GWAS. In reality, even when using sequence data, it is likely that at least a portion of the QTN are not included in the GWAS, for example, because of filtering on allele frequency or imputation accuracy, or because most sequence datasets exclude larger structural variants. Because LD is conserved over shorter distances across breeds than within breeds (De Roos *et al.* 2008), not having the QTN in the dataset and relying on LD between the QTN and other variants nearby may reduce the potential advantage of multi-breed analyses over within-breed analyses.

Overall, we did not find large differences between the results obtained with the meta-analysis and those from multi-breed GWAS, which is in agreement with previously reported results for real milk production traits in dairy cattle (van den Berg *et al.* 2016a; Teissier *et al.* 2018). Hence, a meta-analysis can be a suitable approximation of a multi-breed GWAS when within-breed GWAS summary statistics are available, but not the phenotypes and genotypes required for a multi-breed GWAS. Another advantage of the meta-analysis is the reduced computational demand. The multi-breed GWAS took between 7 and 53 h to complete (running on 20 threads), whereas the time taken for the meta-analysis included 40–90 min for each within-breed GWAS (on 20 threads) and less than a minute for the meta-analysis itself.

COJO is an attractive method to reduce the number of significant variants selected by GWAS and have an estimate of the number of independent signals affecting a quantitative trait (Yang *et al.* 2012). In our simulation study, detection power was insufficient to detect all QTN, which is realistic for quantitative traits where many effects are small and require very large sample sizes for detection (Yengo *et al.* 2022), especially with the lower heritabilities. Consequently, in this situation, COJO tended to underestimate the number of independent signals. Our results also highlighted that the variants selected by COJO are not necessarily the QTN. This can be explained by the extensive LD found in dairy cattle (De Roos *et al.* 2008), resulting in many sequence variants associated with the same QTL region. Differences in imputation accuracy and association with other QTN may result in variants near the QTN being more significant than the QTN. Consequently, COJO might select the more significant variant rather than the true QTN. Hence, when undertaking a GWAS to select potential causal variants for further validation studies, it is likely to be better to select several variants per QTL region rather than restrict the

selection to the most significant variant per peak or variants selected by COJO.

## Conclusions

Our results showed that multi-breed GWAS and meta-analysis generally detect more QTN with improved precision than do within-breed GWAS, particularly if QTN have a similar MAF. However, when allele frequencies differed among breeds, or variants were breed-specific, the advantage of the multi-breed analyses over within-breed GWAS decreased. Regardless of the type of QTN, multi-breed GWAS and meta-analyses performed similarly or better than did within-breed GWAS, demonstrating the benefits of multi-breed GWAS. Additionally, our results obtained with the meta-analysis and multi-breed GWAS were generally not very different; thus, a meta-analysis can be a suitable approximation of a balanced multi-breed GWAS when within-breed GWAS summary statistics are available or reduced computational demand is important.

## Supplementary material

Supplementary material is available [online](#).

## References

- Bellinge RHS, Liberles DA, Iaschi SPA, O'Brien PA, Tay GK (2005) *Myostatin* and its implications on animal breeding: a review. *Animal Genetics* **36**, 1–6. doi:10.1111/j.1365-2052.2004.01229.x
- Bouwman AC, Veerkamp RF (2014) Consequences of splitting whole-genome sequencing effort over multiple breeds on imputation accuracy. *BMC Genetics* **15**, 105. doi:10.1186/s12863-014-0105-8
- Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, Sahana G, Govignon-Gion A, Boitard S, Dolezal M, Pausch H, Brøndum RF, Bowman PJ, Thomsen B, Gulbrandsen B, Lund MS, Servin B, Garrick DJ, Reecy J, Vilkki J, Bagnato A, Wang M, Hoff JL, Schnabel RD, Taylor JF, Vinkhuyzen AAE, Panitz F, Bendixen C, Holm L-E, Gredler B, Hozé C, Boussaha M, Sanchez M-P, Rocha D, Capitan A, Tribout T, Barbat A, Croiseau P, Drögemüller C, Jagannathan V, Vander Jagt C, Crowley JJ, Bieber A, Purfield DC, Berry DP, Emmerling R, Götz K-U, Frischknecht M, Russ I, Sölkner J, Van Tassell CP, Fries R, Stothard P, Veerkamp RF, Boichard D, Goddard ME, Hayes BJ (2018) Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nature Genetics* **50**, 362–367. doi:10.1038/s41588-018-0056-5
- Brøndum RF, Gulbrandsen B, Sahana G, Lund MS, Su G (2014) Strategies for imputation to whole genome sequence using a single or multi-breed reference population in cattle. *BMC Genomics* **15**, 728. doi:10.1186/1471-2164-15-728
- Brøndum RF, Su G, Janss L, Sahana G, Gulbrandsen B, Boichard D, Lund MS (2015) Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. *Journal of Dairy Science* **98**, 4107–4116. doi:10.3168/jds.2014-9005
- Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brøndum RF, Liao X, Djari A, Rodriguez SC, Grohs C, Esquerré D, Bouchez O, Rossignol M-N, Klopp C, Rocha D, Fritz S, Eggen A, Bowman PJ, Coote D, Chamberlain AJ, Anderson C, VanTassell CP, Hulsege I, Goddard ME, Gulbrandsen B, Lund MS, Veerkamp RF, Boichard DA, Fries R, Hayes BJ (2014) Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics* **46**, 858–865. doi:10.1038/ng.3034
- Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh P-R, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis Gçalo R, Fuchsberger C (2016) Next-generation genotype imputation service and methods. *Nature Genetics* **48**, 1284–1287. doi:10.1038/ng.3656
- de Roos APW, Hayes BJ, Spelman RJ, Goddard ME (2008) Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. *Genetics* **179**, 1503–1512. doi:10.1534/genetics.107.084301
- Gautier M, Capitan A, Fritz S, Eggen A, Boichard D, Druet T (2007) Characterization of the DGAT1 K232A and variable number of tandem repeat polymorphisms in French Dairy Cattle. *Journal of Dairy Science* **90**, 2980–2988. doi:10.3168/jds.2006-707
- Hayes BJ, Daetwyler HD (2019) 1000 bull genomes project to map simple and complex genetic traits in cattle: applications and outcomes. *Annual Review of Animal Biosciences* **7**, 89–102. doi:10.1146/annurev-animal-020518-115024
- Jiang J, Ma L, Prakapenka D, VanRaden PM, Cole JB, Da Y (2019) A large-scale genome-wide association study in U.S. Holstein cattle. *Frontiers in Genetics* **10**, 412. doi:10.3389/fgene.2019.00412
- Kemper KE, Hayes BJ, Daetwyler HD, Goddard ME (2015) How old are quantitative trait loci and how widely do they segregate? *Journal of Animal Breeding and Genetics* **132**, 121–134. doi:10.1111/jbg.12152
- Loh P-R, Palamara PF, Price AL (2016) Fast and accurate long-range phasing in a UK Biobank cohort. *Nature Genetics* **48**, 811–816. doi:10.1038/ng.3571
- Marete AG, Gulbrandsen B, Lund MS, Fritz S, Sahana G, Boichard D (2018) A meta-analysis including pre-selected sequence variants associated with seven traits in three French dairy cattle populations. *Frontiers in Genetics* **9**, 522. doi:10.3389/fgene.2018.00522
- Pausch H, MacLeod IM, Fries R, Emmerling R, Bowman PJ, Daetwyler HD, Goddard ME (2017) Evaluation of the accuracy of imputed sequence variant genotypes and their utility for causal variant detection in cattle. *Genetics Selection Evolution* **49**, 24. doi:10.1186/s12711-017-0301-x
- Raven L-A, Cocks BG, Hayes BJ (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
- Raymond B, Bouwman AC, Schrooten C, Houwing-Duistermaat J, Veerkamp RF (2018a) Utility of whole-genome sequence data for across-breed genomic prediction. *Genetics Selection Evolution* **50**, 27. doi:10.1186/s12711-018-0396-8
- Raymond B, Bouwman AC, Wientjes YCJ, Schrooten C, Houwing-Duistermaat J, Veerkamp RF (2018b) Genomic prediction for numerically small breeds, using models with pre-selected and differentially weighted markers. *Genetics Selection Evolution* **50**, 49. doi:10.1186/s12711-018-0419-5
- Sargolzaei M, Chesnais JP, Schenkel FS (2014) A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* **15**, 478. doi:10.1186/1471-2164-15-478
- Teissier M, Sanchez MP, Boussaha M, Barbat A, Hoze C, Robert-Granie C, Croiseau P (2018) Use of meta-analyses and joint analyses to select variants in whole genome sequences for genomic evaluation: an application in milk production of French dairy cattle breeds. *Journal of Dairy Science* **101**, 3126–3139. doi:10.3168/jds.2017-13587
- Van den Berg I, Boichard D, Lund MS (2016a) 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. *Journal of Dairy Science* **99**, 8932–8945. doi:10.3168/jds.2016-11073
- Van den Berg I, Boichard D, Lund MS (2016b) Sequence variants selected from a multi-breed GWAS can improve the reliability of genomic predictions in dairy cattle. *Genetics Selection Evolution* **48**, 83. doi:10.1186/s12711-016-0259-0
- Van den Berg I, Xiang R, Jenko J, Pausch H, Boussaha M, Schrooten C, Tribout T, Gjuvsland AB, Boichard D, Nordbø Ø, Sanchez M-P, Goddard ME (2020) Meta-analysis for milk fat and protein percentage using imputed sequence variant genotypes in 94 321

- cattle from eight cattle breeds. *Genetics Selection Evolution* **52**, 37. doi:10.1186/s12711-020-00556-4
- Van den Berg I, Ho PN, Nguyen TV, Haile-Mariam M, MacLeod IM, Beatson PR, O'Connor E, Pryce JE (2022) GWAS and genomic prediction of milk urea nitrogen in Australian and New Zealand dairy cattle. *Genetics Selection Evolution* **54**, 15. doi:10.1186/s12711-022-00707-9
- VanRaden PM, Tooker ME, O'Connell JR, Cole JB, Bickhart DM (2017) Selecting sequence variants to improve genomic predictions for dairy cattle. *Genetics Selection Evolution* **49**, 32. doi:10.1186/s12711-017-0307-4
- Waller CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191. doi:10.1093/bioinformatics/btq340
- Xiang R, Van Den Berg I, MacLeod IM, Hayes BJ, Prowse-Wilkins CP, Wang M, Bolormaa S, Liu Z, Rochfort SJ, Reich CM, Mason BA, Vander Jagt CJ, Daetwyler HD, Lund MS, Chamberlain AJ, Goddard ME (2019) Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 19398–19408. doi:10.1073/pnas.1904159116
- Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* **88**, 76–82. doi:10.1016/j.ajhg.2010.11.011
- Yang J, Ferreira T, Morris AP, Medland SE, Madden PAF, Heath AC, Martin NG, Montgomery GW, Weedon MN, Loos RJ, Frayling TM, McCarthy MI, Hirschhorn JN, Goddard ME, Visscher PM, Genetic Investigation of Anthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics* **44**, 369–375. doi:10.1038/ng.2213
- Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, Graff M, Eliassen AU, Jiang Y, Raghavan S, *et al.* (2022) A saturated map of common genetic variants associated with human height. *Nature* **610**, 704–712. doi:10.1038/s41586-022-05275-y

**Data availability.** The data used to generate the results in this paper are not publicly available.

**Conflicts of interest.** The authors declare no conflicts of interest.

**Declaration of funding.** This study was undertaken as part of the DairyBio program, which is jointly funded by Dairy Australia (Melbourne, Australia) and Agriculture Victoria (Melbourne, Australia) and The Gardiner Foundation (Melbourne, Australia).

**Acknowledgements.** This study was undertaken as part of the DairyBio program, which is jointly funded by Dairy Australia (Melbourne, Australia) and Agriculture Victoria (Melbourne, Australia) and The Gardiner Foundation (Melbourne, Australia). The authors extend their gratitude to Dr Tuan Nguyen and Dr Bolormaa Sunduimijid for imputation, and the farmers and DataGene (Bundoora, Australia) for access to some genotype data used in this study. We thank partners from the 1000 Bulls genomes project for access to Run8 data for imputation.

#### Author affiliations

<sup>A</sup>Agriculture Victoria, AgriBio, Centre for AgriBioscience, 5 Ring Road, Bundoora, Vic. 3083, Australia.

<sup>B</sup>School of Applied Systems Biology, La Trobe University, 5 Ring Road, Bundoora, Vic. 3083, Australia.