frontiers in Genetics

ORIGINAL RESEARCH
published: 29 April 2021
doi: 10.3389/fgene.2021.642065

Accounting for Population Structure and Phenotypes From Relatives in Association Mapping for Farm Animals: A Simulation Study

Enrico Mancin[1,2]*, *Daniela Lourenco*[2], *Matias Bermann*[2], *Roberto Mantovani*[1] and *Ignacy Misztal*[2]

[1] *Department of Agronomy, Food, Natural Resources, Animals and Environment, University of Padova, Padua, Italy,*
[2] *Department of Animal and Dairy Science, University of Georgia, Athens, GA, United States*

Population structure or genetic relatedness should be considered in genome association studies to avoid spurious association. The most used methods for genome-wide association studies (GWAS) account for population structure but are limited to genotyped individuals with phenotypes. Single-step GWAS (ssGWAS) can use phenotypes from non-genotyped relatives; however, its ability to account for population structure has not been explored. Here we investigate the equivalence among ssGWAS, efficient mixed-model association expedited (EMMAX), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS), and how they differ from the single-SNP analysis without correction for population structure (SSA-NoCor). We used simulated, structured populations that mimicked fish, beef cattle, and dairy cattle populations with 1040, 5525, and 1,400 genotyped individuals, respectively. Larger populations were also simulated that had up to 10-fold more genotyped animals. The genomes were composed by 29 chromosomes, each harboring one QTN, and the number of simulated SNPs was 35,000 for the fish and 65,000 for the beef and dairy cattle populations. Males and females were genotyped in the fish and beef cattle populations, whereas only males had genotypes in the dairy population. Phenotypes for a trait with heritability varying from 0.25 to 0.35 were available in both sexes for the fish population, but only for females in the beef and dairy cattle populations. In the latter, phenotypes of daughters were projected into genotyped sires (i.e., deregressed proofs) before applying EMMAX and SSA-NoCor. Although SSA-NoCor had the largest number of true positive SNPs among the four methods, the number of false negatives was two–fivefold that of true positives. GBLUP-GWAS and EMMAX had a similar number of true positives, which was slightly smaller than in ssGWAS, although the difference was not significant. Additionally, no significant differences were observed when deregressed proofs were used as pseudo-phenotypes in EMMAX compared to daughter phenotypes in ssGWAS for the dairy cattle population. Single-step GWAS accounts for population structure and is a straightforward method for association analysis when only a fraction of the population is genotyped and/or when phenotypes are available on non-genotyped relatives.

Keywords: genome-wide association, genetic relatedness, single-step genomic best linear unbiased prediction, spurious associations, mixed models, single step genome-wide association study

OPEN ACCESS

Edited by:
James Reecy,
Iowa State University, United States

Reviewed by:
Robert John Tempelman,
Michigan State University,
United States
Renata Veroneze,
Universidade Federal de Viçosa, Brazil

****Correspondence:***
Enrico Mancin
enrico.mancin@studenti.unipd.it

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

Received: *15 December 2020*
Accepted: *07 April 2021*
Published: *29 April 2021*

Citation:
Mancin E, Lourenco D,
Bermann M, Mantovani R and
Misztal I (2021) Accounting
for Population Structure
and Phenotypes From Relatives
in Association Mapping for Farm
Animals: A Simulation Study.
Front. Genet. 12:642065.
doi: 10.3389/fgene.2021.642065

INTRODUCTION

Genome-wide association (GWA) aims to identify regions in the genome that are related to diseases or traits of interest (Begum et al., 2012). The method is most often based on statistical tests to determine if a single nucleotide polymorphism (SNP) is statistically associated with the trait, at a given probability value (*p*-value). If the association is significant, the interrogated SNP may be in high linkage disequilibrium (LD) with a causative variant, or the SNP itself may be a common variant that has a large effect on the trait, although having one or a few causative variants and validating them can be difficult (Kennedy et al., 1992). In fact, results from GWA study (GWAS) have confirmed that most of the complex traits in humans (Yang et al., 2011), animals (de Oliveira Silva et al., 2017), and plants (Bian and Holland, 2017) are polygenic. Even in such a case, the GWAS still fulfills the primary goal of helping to better understand the biology of a trait.

The first GWAS was developed to understand the biology of human diseases aiming the prevention (Burton et al., 2007). Although a couple of studies were published a few years before, the study from 2007 is considered the landmark of GWAS because it resulted from a well-designed, large-scale study (Visscher et al., 2012). After that, GWAS was also adopted in livestock and plants. The very first studies were based on single-SNP analysis where each SNP is tested independently (Balding, 2006). However, this approach assumes SNPs are identically and independently distributed, which is only true when a population is comprised of unrelated individuals (Risch and Merikangas, 1996). As populations contain related individuals, not considering population structure or genetic relatedness in GWAS can result in spurious associations (Sul et al., 2018) To resolve the problem with population structure, the use of principal components (PC) to model relationships have been suggested (Price et al., 2006). Still, the level of confounding in GWAS was considerable when 100 PC were fit into the model or when highly related individuals were removed from a human population (Sul et al., 2018).

A well-known approach among animal breeders, the mixed linear models (Henderson, 1975), was then adopted for human GWAS showing to be a reasonable approach to take population structure into account (Kang et al., 2008, 2010). In this method, known as efficient mixed-model association expedited (EMMAX), one SNP is fit in the model as a fixed covariate and, at the same time, a relationship matrix corrects for population structure. However, EMMAX-based methods consider only genotyped individuals with phenotypes. However, only a fraction of individuals in a population are genotyped, particularly in livestock and aquaculture. Because of that, the original mixed linear models were extended to account for genotyped and non-genotyped individuals in prediction analysis (Aguilar et al., 2010; Christensen and Lund, 2010). This method is called single-step genomic best linear unbiased prediction (ssGBLUP) and is widely adopted for genomic predictions in livestock (Legarra et al., 2014; Misztal et al., 2020) and plants (Cappa et al., 2019), and was recently applied to predict polygenic risk score in humans (Truong et al., 2020). The popularity of ssGBLUP is due to the added value of phenotypes for relatives that are not genotyped, and the simplicity when combining information from genotyped and non-genotyped individuals (Legarra et al., 2014).

The usefulness of ssGBLUP to GWAS in a procedure called single-step GWAS (ssGWAS) was subsequently extended (Wang et al., 2012). In this method, SNP effects and variance explained by SNPs are computed simultaneously for all SNPs while accounting pedigree and genomic relationships in addition to all phenotypes available. However, no statistical significance test was available under the ssGWAS framework. Later it has been shown that the statistical test used in EMMAX has a mathematical equivalent that can be used in GBLUP-based methods (Gualdrón Duarte et al., 2014; Bernal Rubio et al., 2016) even though SNPs are considered fixed in the former and random in the latter. This equivalent statistical test was then implemented in ssGWAS (Aguilar et al., 2019) so that *p*-values are computed based on prediction error variance of SNP effects. Because of the mathematical equivalence, results from ssGWAS are expected to be similar to the ones from EMMAX. Here we use different simulated, structured populations (i.e., beef cattle, dairy cattle, and fish) to investigate the equivalence among EMMAX, ssGWAS, and GBLUP-GWAS, and how they differ from single-SNP analysis. We also evaluate whether the population structure is fully considered by the mixed linear models, and when ssGWAS should be the method of choice for association studies in related populations. We demonstrate ssGWAS performs similarly to EMMAX and GBLUP-GWAS when genotyped animals have their own phenotypes or when only progeny phenotypes are available, so deregressed proofs have to be used for EMMAX and GBLUP-GWAS.

MATERIALS AND METHODS

Methods and Computations

Single-SNP Analysis Without Correction for Population Structure (SSA-NoCor)

To estimate allele substitution effect of the i^{th} SNP with SSA-NoCor, the following model was used:

$$y = 1\mu + x_i g_i + e \tag{1}$$

where y is the vector of phenotypes, μ is the mean, x_i is a vector that contains the genotype for the i^{th} SNP for each animal, g_{i} is the i^{th} allele substitution effect, considered as fixed, and $e \sim N(0, I\sigma_e^2)$ is the residual. The estimate of g_{i} $\left(\hat{g}_{\mathrm{i}}\right)$ and its variance $\left(\mathrm{Var}\left(\hat{g}_{\mathrm{i}}\right)\right)$ were obtained by least squares.

Single-SNP Analysis With Correction for Population Structure Using a Genomic Relationship Matrix (EMMAX)

For the EMMAX method, the estimated allele substitution effect and its variance were obtained from the BLUE of the following linear mixed model:

$$y = 1\mu + x_{\mathrm{i}} g_{\mathrm{i}} + \mathrm{Za} + e \tag{2}$$

where Z is a design matrix, $a \sim N(0, G\sigma_{\mathrm{a}}^2)$ is the vector of breeding values (i.e., animal effect), G is the genomic relationship matrix, and the rest of the components were previously defined. The G matrix was calculated as first method following literature (Zhou and Stephens, 2012):

$$G = \frac{1}{k} \sum_{i=1}^{k} (x_{\mathrm{i}} - 1_{\mathrm{n}}\bar{\mathrm{x}}_{\mathrm{i}})(x_{\mathrm{i}} - 1_{n}\bar{\mathrm{x}}_{i})^{T} \tag{3}$$

where x_{i} stands for i^{th} SNP locus column, $\bar{\mathrm{x}}_{\mathrm{i}}$ represented the marker sample mean of i^{th} locus, n and k are the number of genotyped animals and SNPs.

GBLUP Association (GBLUP-GWAS)

For the GBLUP-GWAS, the vector of estimated allele substitution effects $\hat{g}$ was obtained from a linear transformation of the BLUP of a under a GBLUP model:

$$y = 1\mu + Za + e \tag{4}$$

of which the mixed model equations can be represented by:

$$\begin{bmatrix} 1^{'}1 & 1^{'}Z \\ Z^{'}1 & Z^{'}Z + \mathrm{G}^{-1}\frac{\sigma_{\mathrm{e}}^{2}}{\sigma_{\mathrm{a}}^{2}} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} 1^{'}y \\ Z^{'}y \end{bmatrix} \tag{5}$$

In this method, G_0 was estimated using the first method of VanRaden (2008):

$$G_0 = \frac{\mathrm{MM}'}{2\ \sum p_{\mathrm{i}}\,(1 - p_{\mathrm{i}}\,)} \tag{6}$$

where M is a matrix of SNP content centered by twice the current allele frequencies, and p_i is the allele frequency for the i^{th} SNP. Additionally, to avoid singularity problems, the final G was computed as

$$G = \lambda G_0 + \beta I \tag{7}$$

with λ = 0.95 and β = 0.05.

Afterward, the vector of allele substitution effects ($\hat{g}$) was calculated for all SNPs simultaneously (Wang et al., 2012):

$$\hat{g} = \lambda \frac{1}{2\sum pq} M'G^{-1}\hat{a} \tag{8}$$

with $q = 1 - p$.

The variance of SNP effects, which is needed to compute p-values when SNPs are considered random was calculated following Gualdrón Duarte et al. (2014):

$$Var(\hat{g}) = \lambda \frac{1}{2\sum pq} Z'G^{-1}(G\hat{\sigma}_{a}^{2} - C^{22})G^{-1}Z\ \lambda \frac{1}{2\sum pq} \tag{9}$$

where C^{22} is the block of the inverse of the MME corresponding to the animal effect. The p-value for each SNP effect was then computed with the formula (Gualdrón Duarte et al., 2014):

$$p - value_{\mathrm{i}} = 2\left(1 - \Phi\left(\left|\frac{\hat{g}_{\mathrm{i}}}{sd(\hat{g}_{\mathrm{i}})}\right|\right)\right) \tag{10}$$

where $sd(\hat{g}_{\mathrm{i}})$ is the standard error of the SNP effect or simply $sd(\hat{g}_{\mathrm{i}}) = \sqrt{Var(\hat{g}_{\mathrm{i}})}$; $\Phi(\cdot)$ is the cumulative density function (CDF) of the standard normal distribution. For a justification of using $sd(\hat{g}_{\mathrm{i}}) = \sqrt{Var(\hat{g}_{\mathrm{i}})}$ in the denominator instead of $\sqrt{Var(g_{\mathrm{i}}) - Var(\hat{g}_{\mathrm{i}})}$ (Gualdrón Duarte et al., 2014; Bernal Rubio et al., 2016).

Single-Step GBLUP Association (ssGWAS)

This method differs from GBLUP-GWAS in the sense that all animals in the pedigree can be used, not only genotyped animals with phenotypes. Therefore, G^{-1} is replaced by H^{-1} in (5), and the latter combines pedigree and genomic relationships (Aguilar et al., 2010):

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix} \tag{11}$$

where A^{-1} and A_{22}^{-1} are the inverses of the pedigree relationship matrix for all animals and only genotyped animals, respectively. Pedigree and genomic relationships have different genetic base because allele frequencies from the current genotyped population are used to center G. Therefore, G in ssGWAS is adjusted so the average diagonal and off-diagonal matches the averages of A_{22}. Because of this adjustment, (8) and (9) were modified to:

$$\hat{g} = \lambda\delta \frac{1}{2\sum pq} M'G^{-1}\hat{a}_{22} \tag{12}$$

and

$$Var(\hat{g}) = \lambda\delta \frac{1}{2\sum pq} Z'G^{-1}(G\hat{\sigma}_{\mathrm{a}}^{2} - C^{22})G^{-1}Z\ \lambda\delta \frac{1}{2\sum pq} \tag{13}$$

where $\hat{a}_{22}$ is a vector of genomic estimated breeding values (GEBV) for genotyped animals; δ accounts for the difference in genetic base between the pedigree and genomic relationship matrices, and was calculated as in Vitezica et al. (2011):

$$\delta = 1 - \frac{0.5}{n^2}\Big(\sum_{\mathrm{i}}\sum_{\mathrm{j}} A_{22(\mathrm{i,j})} - \sum_{\mathrm{i}}\sum_{\mathrm{j}} G_{\mathrm{i,j}}\Big) \tag{14}$$

with n the number of genotyped animals. After the modification, p-values in ssGWAS were obtained as in (10) as previously suggested (Aguilar et al., 2019).

Note that the dimension of ssGWAS system of equations is greater than the dimension of GBLUP-GWAS because of the inclusion of non-genotyped animals.

Computations

The allele substitution effects were estimated with different software: (i) SSA-NoCor solutions were computed with GASTON R-package (Dandine-Roulland and Perdry, 2018), (ii) EMMAX solutions were obtained with GEMMA software (Zhou and Stephens, 2012), (iii) GBLUP-GWAS and (iv) ssGWAS were computed using the BLUPF90 software suite (Misztal et al., 2014b).

Overview of Data Simulation

We used different simulated datasets to investigate the equivalence between EMMAX and ssGWAS, and to explore the usefulness of each method compared to SSA-NoCor and

GBLUP-GWAS. Small and large fish, beef cattle, and dairy cattle populations were simulated using QMSim (Sargolzaei and Schenkel, 2009), with five replicates for each. Because we wanted to test the efficiency of the methods in accounting for population structure, strong assumptions were used to generate extreme scenarios. The general parameters for each population such as the number of genotyped animals, effective population size, type of trait, and heritability are reported in **Table 1**. The linkage disequilibrium generated in each small and large populations are in **Figures 1**, **2**.

Fish Population

The historical population began with 5,000 animals and decreased to 3,100 after 200 non-overlapping generations, that were carried out to generate linkage disequilibrium (LD) and mutation-drift equilibrium. The proportion of males in the historical population was 32%. Aiming to mimic the beginning of a fish breeding scheme, a recent population was created by randomly selecting 20 sires and 20 dams. The recent population was subject to random selection, culling, and mating for five non-overlapping generations (Garcia et al., 2018). In every generation, each female had 2 offspring. After the fifth generation, a new line was created by randomly mating 20 males and 20 females from generation 5. For this line, the litter size was set to 100 offspring per dam, and only one generation was created. The effective population size was 40. Phenotypes and pedigree of the animals in the new line, together with their parents, were considered for the association analysis. A larger fish population was simulated that had 10 times more animals (**Table 1**).

The genome was composed of 29 chromosomes with a length of approximately 100 cM each, 35,000 evenly spaced SNPs, and one QTN per chromosome. Each QTN was placed in the middle

TABLE 1 | General parameters for the simulated populations[1,2].

Simulated populations	Number of records	Population size	Number of genotyped animals	Number of genotyped animals with records[1]	Trait type	Heritability
Small populations						
Fish	2040	2040	1040	1040	Both sex	0.25
Beef	5088	10,000	5525	3010	Sex-limited	0.30
Dairy	70,000 (0.38)	140,000	1438	1438	Sex-limited	0.35
Large populations						
Fish	20,400	20,400	10,400	10,400	Both sex	0.25
Beef	25,000	50,000	14,250	8985	Sex-limited	0.30
Dairy	250,000 (0.37)	500,000	14,000	14,000	Sex-limited	0.35

[1] *For the dairy population, the records are represented as deregressed proofs. The average reliability of deregressed proofs is reported inside parenthesis.*
[2] *For beef populations, association studies with SSA-NoCor, EMMAX, and GBLUP-GWAS were performed only with genotypes for the sex that had records.*

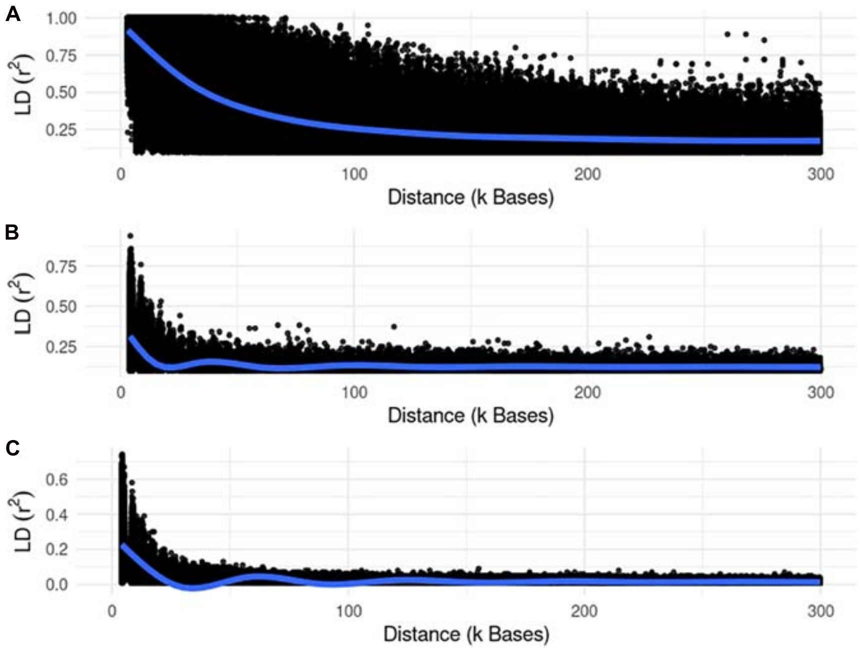

FIGURE 1 | Linkage disequilibrium (r2) decay for the small populations of fish **(A)**, beef **(B)**, and dairy cattle **(C)**.

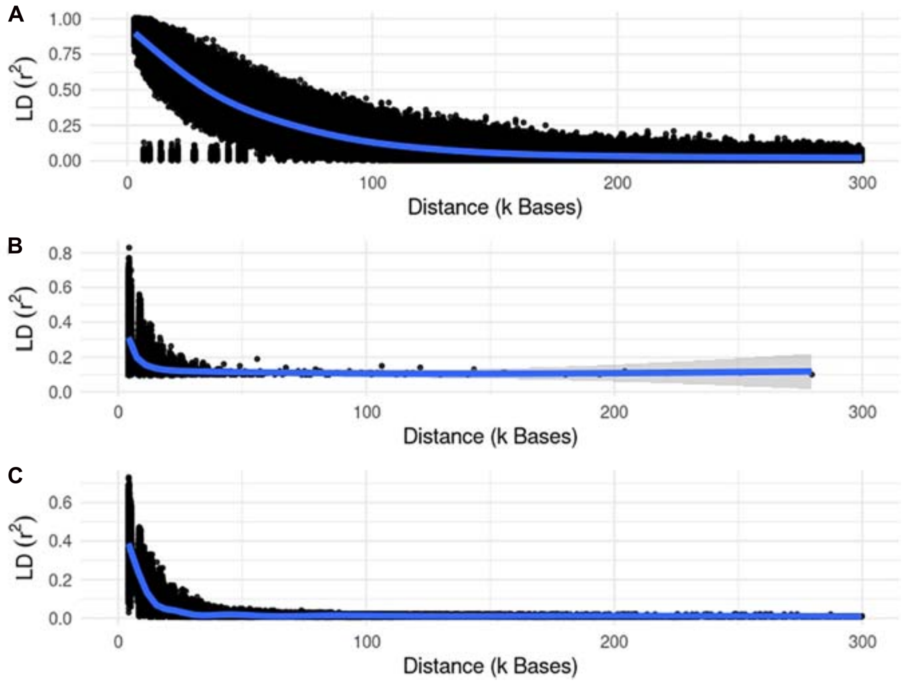

FIGURE 2 | Linkage disequilibrium (r2) decay for the large populations of fish **(A)**, beef **(B)**, and dairy cattle **(C)**.

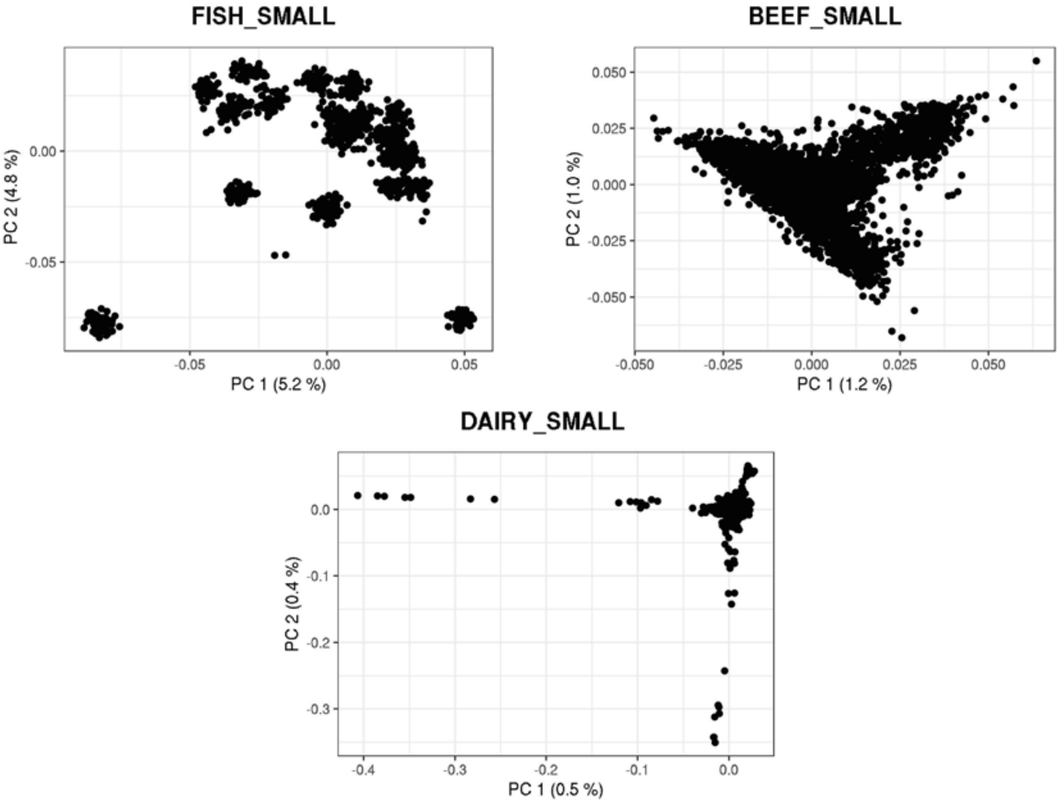

FIGURE 3 | First and second principal components of the genomic relationship matrices for the small populations of fish, beef, and dairy cattle.

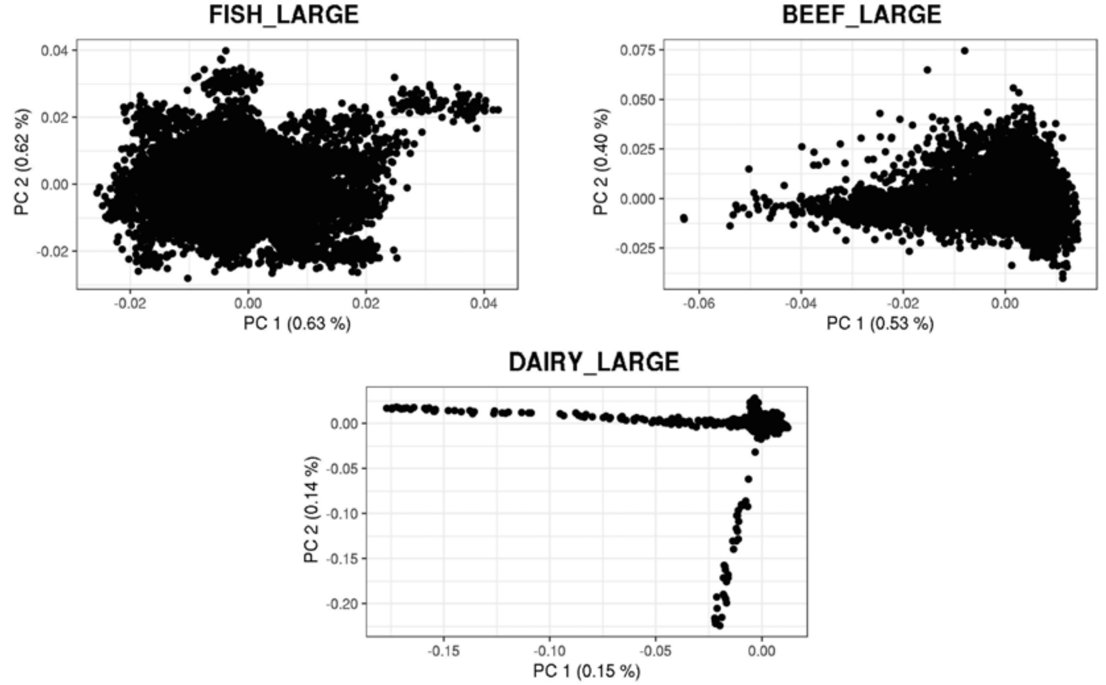

FIGURE 4 | First and second principal components of the genomic relationship matrices for the large populations of fish, beef, and dairy cattle.

of its respective chromosome. Although this number of QTN is not the reality of most of the traits of interest (i.e., complex traits), this assumption was made to facilitates the QTN discovery in the association analysis. Altogether, the QTNs accounted for the total of the genetic variation and their effects were assumed to follow a gamma distribution with shape parameter equal to 0.40. The allele frequencies for SNPs and QTNs in the first historical generation were 0.5, and a recurrent mutation rate of 2.5e-5 per locus per generation was assumed. A single trait with heritability of 0.25 was simulated, and a single phenotype per animal was obtained by adding an overall mean of 1.0, the sum of the QTN effects, and a residual effect.

Beef Cattle Population

In this dataset, the historical population began with 1,000 animals and steadily increased to 50,000 after 1,000 generations of random mating. Then, a decrease in number of individuals followed for another 1,000 generations. After 2,000 generations, the historical population was composed by 23,000 animals, of which 3,000 were males. The recent beef cattle population was created by randomly selecting 10,000 dams and 200 sires, allowing them to mate randomly for five discrete generations. Afterward, five groups of 10 sires and 500 dams each were selected based on EBV to create five different lines. With the aim of maximizing the difference between the five lines, selection based on EBV was used in each of them. Finally, 10 sires and 500 dams from each of the five lines were pooled in one single line and underwent random mating for five generations. This process was designed to create an intricate population structure. For the present population, a sex-limited trait was simulated so that only females had a phenotype for a trait with heritability of 0.30. A larger beef cattle population was simulated that had 5 times more animals (**Table 1**).

Genotypes were simulated for males and females from the last generation of the population and their parents (n_{beef} = 5525). The parameters to simulate the beef cattle genome were the same as in the fish population except for the number of SNPs, which for the beef cattle population was equal to 65,000.

Dairy Cattle Population

The parameters for the simulation of the dairy historical population were the same as those in beef cattle. A total of 1,000 sires and 20,000 were chosen as founders of the recent population. This population was subject to selection based on estimated breeding values (EBV) for 10 generations and mating design based on inbreeding (Sonesson and Meuwissen, 2000). In this simulation, the average Ne was 90. The Ne was calculated as the change in inbreeding (ΔF) from one generation to the next using the following formula (Falconer and Mackay, 1996):

$$\Delta F = \frac{F_{\mathrm{n}} - F_{\mathrm{n-1}}}{1 - F_{\mathrm{n-1}}} \tag{15}$$

$$Ne = \frac{1}{2\Delta F} \tag{16}$$

where F_{n} is the inbreeding in the *nth* generation.

All the parameters for the genome simulation were similar to the ones in the beef cattle population. The only difference was the genotyping strategy that included only sires of the seventh generation (n_{dairy} = 1400). Phenotypes for a trait of heritability

0.35 were available only on females (**Table 1**). On average, each genotyped sire in this population had 10 daughters with records. A larger dairy cattle population was simulated that had 3.5 times more animals and 10 times more genotyped bulls (**Table 1**).

Deregressed Proofs (DRP)

One requirement in association analysis is that individuals should have both genotypes and phenotypes. In some livestock populations, genotypes may be available for males and phenotypes for females (e.g., milk production in dairy cattle). In such a case, DRP are needed as an input for SSA-NoCor, EMMAX, or GBLUP-GWAS. The DRP are projections of female phenotypes into their relatives' genotypes. Because sex-limited traits were simulated for the beef and dairy cattle populations, DRP were computed for sires in both populations following (VanRaden and Sullivan, 2010; Wiggans et al., 2011).

$$DRP = PA + \frac{EBV - PA}{DE_{\mathrm{prog}}/(DE_{\mathrm{prog}} + DE_{\mathrm{PA}} + 1)} \qquad (17)$$

Where PA is parent average; $DE_{\mathrm{prog}} = \left[\frac{EBV_{\mathrm{rel}}}{(1-EBV_{\mathrm{rel}})}\right] - DE_{\mathrm{PA}}$ and is the daughter equivalent from progeny information; and $DE_{\mathrm{PA}} = \frac{PA_{\mathrm{rel}}}{(1-PA_{\mathrm{rel}})}$ is the daughter equivalent from PA. The EBV_{rel} and PA_{rel} are reliabilities of parent average and EBV, respectively. All EBV, PA, and reliabilities used in the DRP formula were computed using the BLUPF90 software suite (Misztal et al., 2014b). The DRP were used for the association analysis of dairy cattle datasets under SSA-NoCor, EMMAX, and GBLUP-GWAS. As ssGWAS uses all phenotypes, genotypes, and pedigree information available, it does not rely on DRP.

Quality Control Prior to the Association Analysis

Quality control of genomic data removed monomorphic SNPs, SNPs with minor allele frequency (MAF) lower than 0.05, and with deviation between observed and expected allele frequencies greater than 0.15. After quality control, and average of 35,000, 58,000, and 58,000 SNPs were kept for the analysis in the fish, beef cattle, and dairy cattle population, respectively.

Significance and Concordance Tests

A single SNP was considered significantly associated with the considered trait when its p-value was smaller than a certain significance level, which was 0.05 with a Bonferroni correction for multiple testing, i.e., $\frac{0.05}{n}$ where n is the number of SNPs in the corresponding simulated population. Additionally, true positive (TP) and false positive (FP) rates were computed for each scenario using a window size of ±2 cM, which is equivalent to 20–30 markers (Toosi et al., 2018).

RESULTS AND DISCUSION

Population Structure

Figures 3, **4** show plots with the first (PC1) and second (PC2) principal components of G for small and large populations, respectively. PC1 and PC2 represent the two largest sources of variation in the data, and are often used to investigate population structure, which was deemed important in our study. The level of population stratification differed among the simulated populations because of the different selection and mating strategies. For the small populations, distinct family groups (full-sibs) were observed for the fish population, with variable size and impact on the model (extreme clusters farther from PC1 = PC2 = 0); however, the PC scores were of small magnitude. No distinct clustering was detected in the beef cattle population, although a level of variability was observed. Overall, it was not possible to discriminate different groups of individuals but some of them appear more genetically different than others. In fact, animals belonging to five different lines were randomly mated for ten generations, and only genotypes for animals in the 7th generation were retained. Therefore, less genetic distance among animals was created. Still, the largest graphical distance was observed in the dairy cattle population, with a large cluster centered in zero and a few animals genetically distant from the main cluster; however, the amount of variation explained by PC1 and PC2 was small. Possibly, this pattern resulted from non-overlapping generations that created extra genetic distance among some animals. These distant animals were sires that had EBV departing from the population mean. For the large populations, no clustering was observed for the fish and beef populations, whereas the same pattern was observed for dairy cattle in both populations.

Association Analysis

Manhattan plots with p-values for the fish, beef cattle, and dairy cattle populations using SSA-NoCor, EMMAX, and ssGWAS, and GBLUP-GWAS are in **Figures 5–10**. Although one QTN was simulated in each chromosome, the signals were not equally strong because of the assumption of a Gamma distribution, and the selection that the populations underwent. Overall, selection caused fixation for 5–6% of the QTN. To better access the information in the Manhattan plots, the average number of true and false positive SNPs were computed and placed in **Table 2**. Although the number of TP and FP differed among EMMAX, ssGWAS, and GBLUP-GWAS, the differences were not statistically significant (p-value > 0.05).

For all the simulated populations, the greatest number of false positive SNPs was observed for SSA-NoCor. These results agree with those from previous studies (e.g., Yang et al., 2014), which showed that the number of false positives drastically decreased when correcting for population structure. For the small populations, the number of false positive SNPs in SSA-NoCor for the dairy dataset was the smallest one compared to the other simulated populations, whereas for the large populations, the smallest number of false positive SNPs occurred in the beef population. In both cases, the fish population had the greatest number of false positive signals. Since this population had a strong structure (e.g., several separate clusters), it can be concluded that the population structure is related to the number of false positive signals captured by SSA-NoCor. False positive associations capture SNPs that relate to the genetic differences between sub-populations and also with the trait considered (Sul et al., 2018). These spurious signals can also

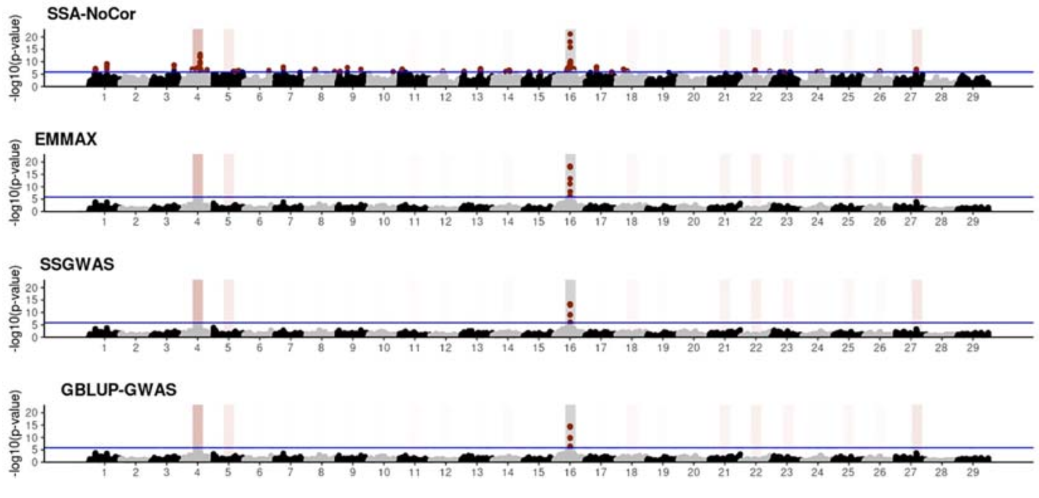

FIGURE 5 | Manhattan plots for the small population of fish using single-SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). Significant SNPs are indicated in red, whereas vertical bars indicate the position of the simulated quantitative trait nucleotide (QTN). The darker the vertical bar, the stronger QTN effect. The blue horizontal line corresponds to the rejection threshold based on a significance level of 0.05 with a Bonferroni correction for multiple testing.

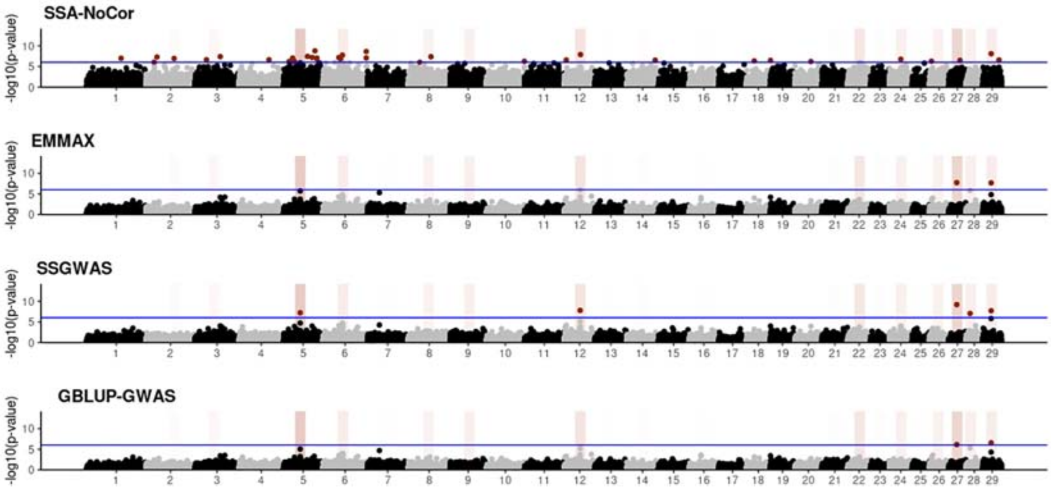

FIGURE 6 | Manhattan plots for the small population of beef cattle using single-SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). Significant SNPs are indicated in red, whereas vertical bars indicate the position of the simulated quantitative trait nucleotide (QTN). The darker the vertical bar, the stronger QTN effect. The blue horizontal line corresponds to the rejection threshold based on a significance level of 0.05 with a Bonferroni correction for multiple testing.

be interpreted as a wrong prior assumption of marker effects in the SSA-NoCor model. In such a model, markers are considered independently distributed, which implicitly means linkage disequilibrium among SNPs is neglected (Finno et al., 2014; Sul et al., 2018). The effect of population structure is even more evident when small sample size and high-density panels are used in association analyses (Finno et al., 2014). Furthermore, when traits are polygenic or have low heritability, signals deriving from population structure can completely override those deriving from true QTNs (Atwell et al., 2010; Toosi et al., 2018).

In terms of true positives, two situations were observed. First, SSA-NoCor detected significantly more true positives than the other methods in the fish and large dairy cattle simulated populations. However, the number of false positives was sometimes threefold greater than that of true positives. The identification of false and true positives is straightforward in simulated data but not in real data. Second, no significant differences were observed among methods for the beef and the small dairy population. Based on the results from the dairy population, it can be concluded that the use of DRP for bulls in EMMAX and GBLUP-GWAS, compared to the raw phenotypes in ssGWAS, did not promote a loss in GWA resolution. The loss in the ability to correctly detect QTNs when using DRP is expected in complex models when the estimation of fixed and

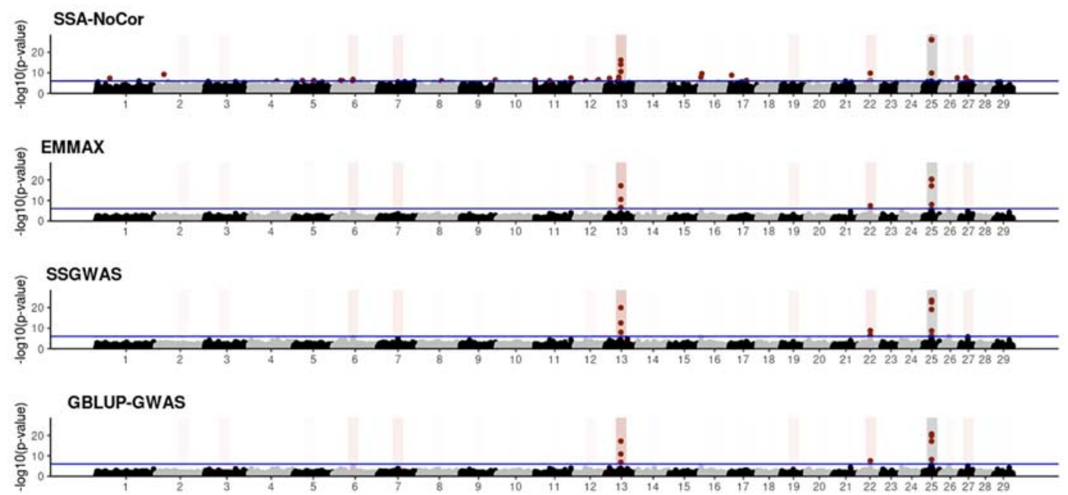

FIGURE 7 | Manhattan plots for the small population of dairy cattle when sires had an average of ten daughters. The association methods used were single-SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). Significant SNPs are indicated in red, whereas vertical bars indicate the position of the simulated quantitative trait nucleotide (QTN). The darker the vertical bar, the stronger QTN effect. The blue horizontal line corresponds to the rejection threshold based on a significance level of 0.05 with a Bonferroni correction for multiple testing.

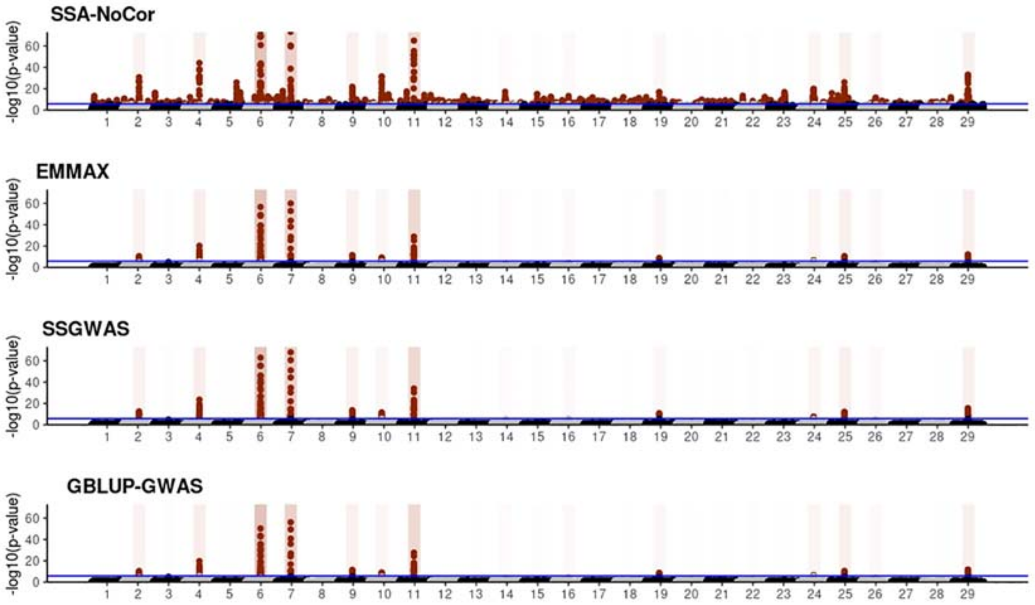

FIGURE 8 | Manhattan plots for the large population of fish using single-SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). Significant SNPs are indicated in red, whereas vertical bars indicate the position of the simulated quantitative trait nucleotide (QTN). The darker the vertical bar, the stronger QTN effect. The blue horizontal line corresponds to the rejection threshold based on a significance level of 0.05 with a Bonferroni correction for multiple testing.

random effects is not very accurate. According to Aguilar et al. (2019), information can be lost in the deregression process, which may result in spurious signals in GWA. Although DRP were used in the dairy population, no significant differences in TP and FP were observed between EMMAX and ssGWAS because the model was simple and included only a general mean as fixed effect and the additive genetic as random.

The methods in our study were used as binary classifiers when trying to identify true and false positives. The quality of a binary classifier can be evaluated from the degree of randomness of the decisions of that classifier. A perfect classifier is not random, whereas the worst classifier would determine whether a signal is true or false with a probability equal to 0.5 (Agresti, 2013). To compare the methods in our

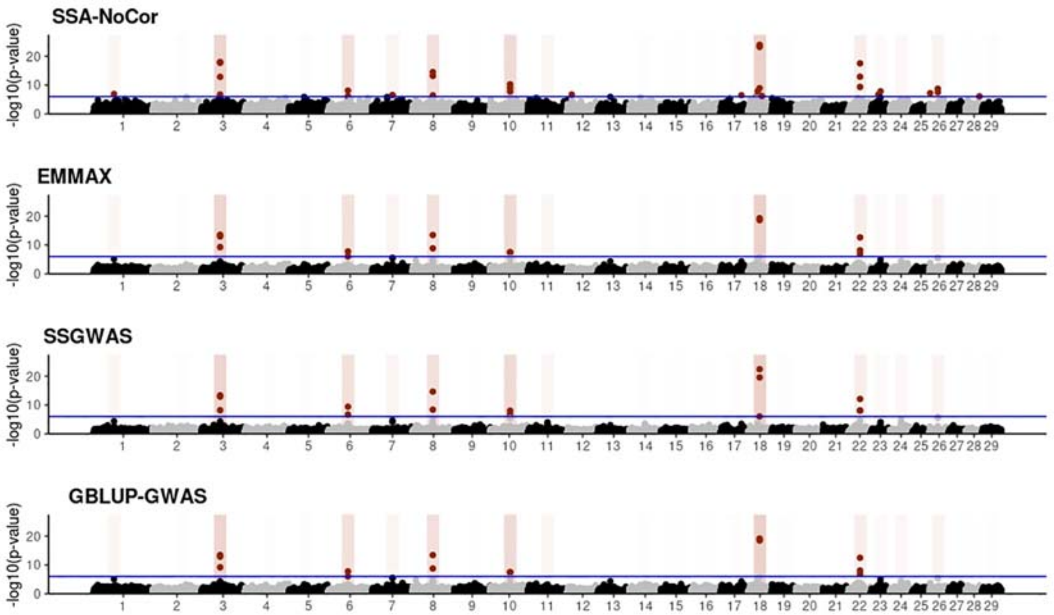

FIGURE 9 | Manhattan plots for the large population of beef cattle using single-SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). Significant SNPs are indicated in red, whereas vertical bars indicate the position of the simulated quantitative trait nucleotide (QTN). The darker the vertical bar, the stronger QTN effect. The blue horizontal line corresponds to the rejection threshold based on a significance level of 0.05 with a Bonferroni correction for multiple testing.

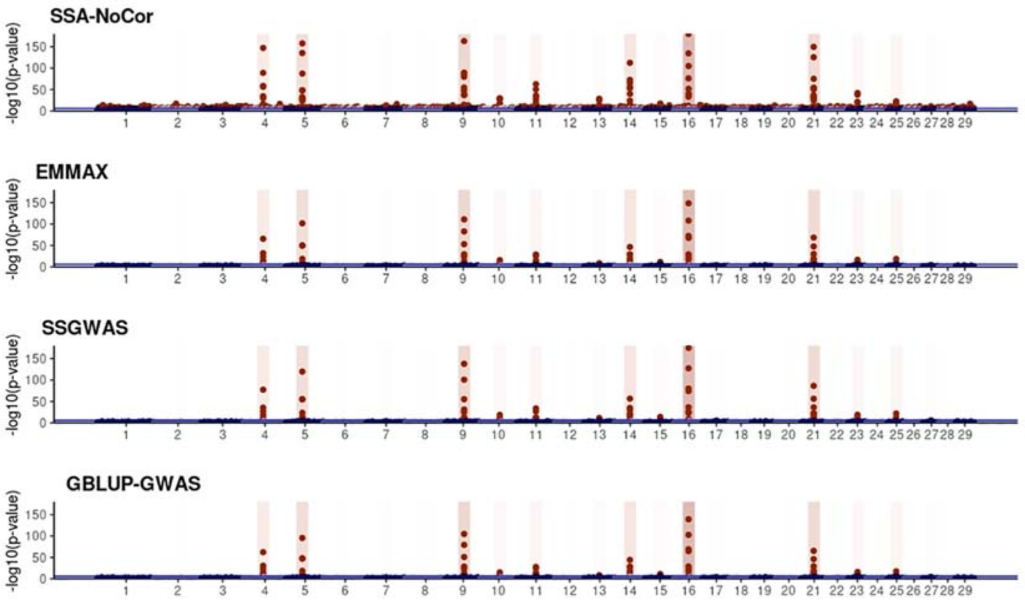

FIGURE 10 | Manhattan plots for the large population of dairy cattle when sires had an average of ten daughters. The association methods used were single-SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). Significant SNPs are indicated in red, whereas vertical bars indicate the position of the simulated quantitative trait nucleotide (QTN). The darker the vertical bar, the stronger QTN effect. The blue horizontal line corresponds to the rejection threshold based on a significance level of 0.05 with a Bonferroni correction for multiple testing.

study as binary classifiers, Receiver Operating Characteristic (ROC) curves were provided for each simulated population in **Figures 11**, **12**. Among the plots, it can be observed that the curve corresponding to SSA-NoCor is the lowest one. Therefore, as a binary classifier, SSA-NoCor performs worse than the rest of the methods. The classifier ability of the models

TABLE 2 | Average (SD) number of true positive (TP) and false positive (FP) SNPs for all the simulated populations[1].

Population	Association	SSA_NoCor	EMMAX	ssGWAS	GBLUP-GWAS
Small populations					
Fish	TP	85.20 (0.47)^{a}	20.00 (0.31)^{b}	20.60 (0.35)^{b}	20.00 (0.31)^{b}
	FP	253.6 (2.50)^{a}	3.80 (1.01)^{b}	5.20 (1.18)^{b}	2.80 (0.96)^{b}
Beef	TP	15.80 (0.68)^{a}	5.60 (0.73)^{a}	7.60 (0.50)^{a}	5.60 (0.73)^{a}
	FP	68.60 (0.78)	0.00	0.20 (2.24)	0.00
Dairy	TP	13.40 (0.20)^{a}	8.60 (0.60)^{a}	12.60 (0.12)^{a}	9.00 (0.55)^{a}
	FP	24.20 (0.32)	0.00	0.20 (2.24)	0.20 (2.24)
Large populations					
Fish	TP	321.40 (72.52)^{a}	106.20 (25.41)^{b}	133.80 (37.93)^{b}	104.20 (25.33)^{b}
	FP	1068 (241.67)	0.00	0.00	0.00
Beef	TP	24.60 (3.29)^{a}	17.20 (5.54)^{a}	21.00 (5.34)^{a}	17.00 (5.48)^{a}
	FP	4.40 (2.19)^{a}	0.40 (0.55)^{b}	0.20 (0.45)^{b}	0.40 (0.55)^{b}
Dairy	TP	151.60 (5.98)^{a}	55.20 (3.7)^{b}	61.60 (8.65)^{b}	52.40 (5.94)^{b}
	FP	853.80 (39.32)	0.00	0.00	0.00

[1] Statistical differences were not calculated when the SD of at least one method within a group was null.
The superscripts within a row, group the results that did not differ significantly.

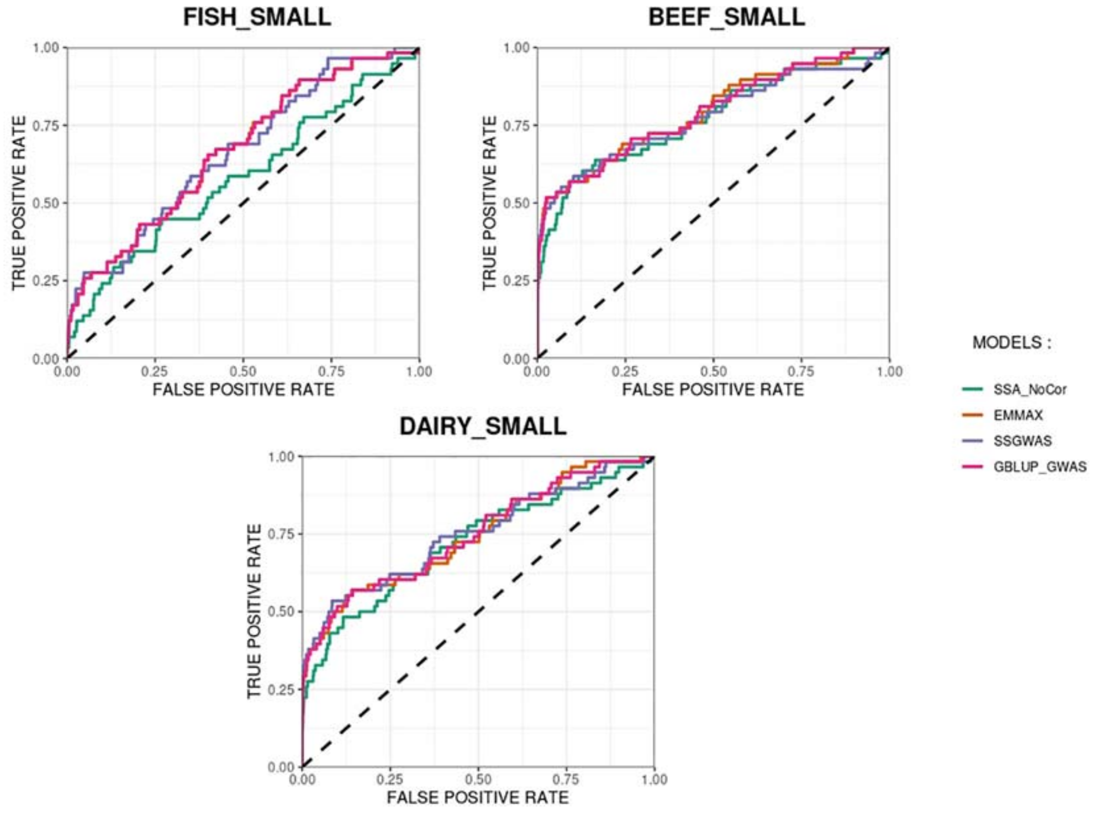

FIGURE 11 | Receiver operating characteristic (ROC) curves for GWAS results for the small populations of fish,beef cattle, dairy cattle with ten daughters per sire. The association methods used were single-SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). The dashed line has slope equal to one and null intercept.

improved in the large populations, but SSA-NoCor still had poorer performance compared to other methods. The fact large data improves the resolution of GWAS is well documented in the literature.

Overall, we observed the size of the populations (e.g., small and large) did not change the outcome of our study, and we confirmed, using simulated populations with intricate structure, that EMMAX, GBLUP-GWAS, and ssGBLUP account

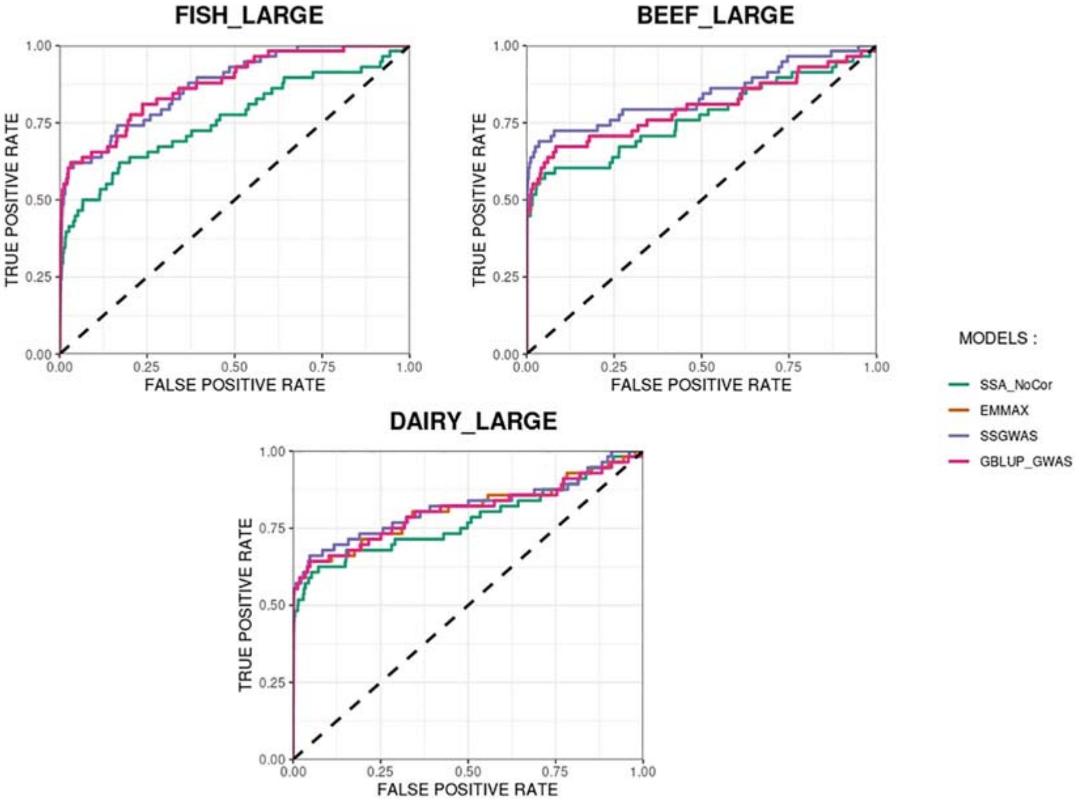

FIGURE 12 | Receiver operating characteristic (ROC) curves for GWAS results for the large populations of fish, beef cattle, dairy cattle with ten daughters per sire. The association methods used were single-SNP analysis without correction for population structure (SSA-NoCor), efficient mixed-model association expedited (EMMAX), single-step GWAS (ssGWAS), and genomic best linear unbiased prediction GWAS (GBLUP-GWAS). The dashed line has slope equal to one and null intercept.

for population structure. The equivalence between p-values obtained in EMMAX and GBLUP-GWAS has been analytically demonstrated (Bernal Rubio et al., 2016), although the former considers SNPs as fixed effects and the latter as random. Lu et al. (2018) extended this idea to single-step and implemented it with the addition of p-values for ssGWAS in the BLUPF90 software suite (Misztal et al., 2014b; Aguilar et al., 2019). This methodology was successfully applied to a beef cattle population with almost 2 million animals in the pedigree, 1 million birth weight records, and a little over 1,400 genotyped sires (Aguilar et al., 2019). In our study, we confirmed that ssGWAS can account for population structure as EMMAX or GBLUP-GWAS.

Recently, single-step was applied for predicting polygenic risk score in humans using phenotypes from related individuals that were not genotyped (Truong et al., 2020). In this study, authors observed an increase in prediction accuracy when raw phenotypes of non-genotyped relatives were included in the model, which is only possible with single-step method. As the number of genotyped individuals in (Legarra et al., 2009; Truong et al., 2020) was 288k, the authors complained about the computing cost of single-step, which is mainly due to the inverse of G. An efficient algorithm to compute G^{-1} without having to directly invert G –the Algorithm for Proven and Young (APY)– is also available (Misztal et al., 2014a). With the APY algorithm, animals are designated as core or non-core, and recursions are done on core animals, whereas predictions for non-core animals are functions of the information for core animals. This is possible because of the assumption that core animals carry all the information about the independent chromosome segments segregating in the population (Misztal, 2016). In addition, it was found (Pocrnic et al., 2016) that the number of largest eigenvalues explaining 98% of the variance in G approaches the number of independent chromosome segments (Stam, 1980) and can be used as the number of core animals in APY. This algorithm enables the computation of genomic predictions for millions of genotyped individuals with much less memory usage and computing time. Indeed, a successful computation of genomic predictions for 13.5 million animals in the pedigree, of which 2.3 million were genotyped, using the BLUPF90 software suite has recently been shown to be feasible (Tsuruta et al., 2021). Although the computation of genomic predictions (GEBV), SNP effects, and variance explained by SNPs can be done efficiently with APY in ssGBLUP, the same does not apply to the computation of p-values in ssGWAS. This is because the formula for p-values (10) relies on the standard error of SNP effects (i.e., square root of prediction error variance), which is currently obtained based on the prediction error variance of GEBV. The latter requires the inverse of the left-hand-side of the single-step mixed model equations, and the computation of inverses of large matrices is extremely expensive. Therefore, the use of ssGWAS may be limited to samples of about 20k genotyped individuals, given the number of total animals in the pedigree is less than 500k. Approximating the prediction error variance of GEBV or SNPs directly may be a way to overcome this limitation, and research on the issue is currently undergoing.

CONCLUSION

Genome-wide association studies in related populations require the correction for population structure to avoid false positive statistical associations between SNPs and trait phenotypes. Several classes of mixed linear models as EMMAX, GBLUP-GWAS, and ssGWAS can take care of this issue by fitting a random effect whose covariance matrix is proportional to a relationship matrix. We demonstrate the three methods did not significantly differ across association studies in several simulated populations, regardless if deregressed proofs or phenotypes from non-genotyped animals are used in the statistical analysis. Further studies are needed to investigate the repeatability of those results in real populations under complex models. Single-step GWAS accounts for population structure as EMMAX or GBLUP-GWAS and allows for the inclusion of phenotypes from non-genotyped relatives.

DATA AVAILABILITY STATEMENT

The original contributions generated for this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

EM performed the analysis and simulation and writing. DL supervised all the analysis, design the global structure of the simulations, and writing. MB revised the manuscript and helpful comments and idea. RM revised the manuscript. IM head of the project. All authors contributed to the article and approved the submitted version.

FUNDING

This study was partially funded by Agriculture and Food Research Initiative Competitive Grant No. 2020-67015-31030 from the United States Department of Agriculture's National Institute of Food and Agriculture.

ACKNOWLEDGMENTS

The authors would like to thank Andres Legarra for his helpful comments.

REFERENCES

Agresti, A. (2013). *Categorical Data Analysis*, 3rd Edn. Hoboken, NJ: John Wiley & Sons Inc.

Aguilar, I., Legarra, A., Cardoso, F., Masuda, Y., Lourenco, D., and Misztal, I. (2019). Frequentist p-values for large scale single step genome wide association, with an application to birth weight in American Angus cattle. *Genet. Select. Evol.* 51:28. doi: 10.1186/s12711-019-0469-3

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (2010). Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93, 743–752. doi: 10.3168/jds.2009-2730

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631. doi: 10.21958/study:1

Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nat. Rev. Genet.* 7, 781–791. doi: 10.1038/nrg1916

Begum, F., Ghosh, D., Tseng, G. C., and Feingold, E. (2012). Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Res.* 40, 3777–3784. doi: 10.1093/nar/gkr1255

Bernal Rubio, Y. L., Gualdrõn Duarte, J. L., Bates, R. O., Ernst, C. W., Nonneman, D., Rohrer, G. A., et al. (2016). Meta-analysis of genome-wide association from genomic prediction models. *Anim. Genet.* 47, 36–48. doi: 10.1111/age.1 2378

Bian, Y., and Holland, J. B. (2017). Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. *Heredity* 118, 585–593. doi: 10.1038/hdy.2017.4

Burton, P. R., Clayton, D. G., Cardon, L. R., Craddock, N., Deloukas, P., Duncanson, A., et al. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678. doi: 10.1038/ng.2007.17

Cappa, E. P., de Lima, B. M., da Silva-Junior, O. B., Garcia, C. C., Mansfield, S. D., and Grattapaglia, D. (2019). Improving genomic prediction of growth and wood traits in Eucalyptus using phenotypes from non-genotyped trees by single-step GBLUP. *Plant Sci.* 284, 9–15. doi: 10.1016/j.plantsci.2019.03.017

Christensen, O., and Lund, M. (2010). Genomic relationship matrix when some animals are not genotyped. *Genet. Select. Evol.* 42, 1–8. doi: 10.1186/1297-9686-42-2

Dandine-Roulland, C., and Perdry, H. (2018). "Manipulation of genetic data (SNPs). Computation of GRM and dominance matrix, LD, heritability with efficient algorithms for linear mixed model (AIREML)," in *46th European Mathematical Genetics Meeting (EMGM) 2018,* Cagliari.

de Oliveira Silva, R., Bonvino Stafuzza, N., de Oliveira Fragomeni, B., de Camargo, G., Matos Ceacero, T., Noely dos Santos Gonçalves Cyrillo, J., et al. (2017). Genome-wide association study for carcass traits in an experimental nelore cattle population. *PLoS One* 12:e0169860. doi: 10.1371/journal.pone.0169860

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. England: Longman Essex.

Finno, C. J., Aleman, M., Higgins, R. J., Madigan, J. E., and Bannasch, D. L. (2014). Risk of false positive genetic associations in complex traits with underlying population structure: a case study. *Vet. J.* 202, 543–549. doi: 10.1016/j.tvjl.2014. 09.013

Garcia, A. L. S., Bosworth, B., Waldbieser, G., Misztal, I., Tsuruta, S., and Lourenco, D. A. L. (2018). Development of genomic predictions for harvest and carcass weight in channel catfish 06 Biological Sciences 0604 Genetics. *Genet. Select. Evol.* 50, 1–12. doi: 10.1186/s12711-018-0435-5

Gualdrón Duarte, J. L., Cantet, R. J. C., Bates, R. O., Ernst, C. W., Raney, N. E., and Steibel, J. P. (2014). Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. *BMC Bioinformatics* 15:246. doi: 10.1186/1471-2105-15-246

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423–447. doi: 10.2307/2529430

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., et al. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42, 348–354. doi: 10.1038/ng.548

Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., et al. (2008). Efficient control of population structure in model organism association mapping. *Genetics* 178, 1709–1723. doi: 10.1534/genetics.107.080101

Kennedy, B. W., Quinton, M., and van Arendonk, J. A. (1992). Estimation of effects of single genes on quantitative traits. *J. Anim. Sci.* 70, 2000–2012. doi: 10.2527/1992.7072000x

Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92, 4656–4663. doi: 10.3168/jds.2009-2061

Legarra, A., Christensen, O. F., Aguilar, I., and Misztal, I. (2014). Single Step, a general approach for genomic selection. *Livestock Sci.* 166, 54–65. doi: 10.1016/j.livsci.2014.04.029

Lu, Y., Vandehaar, M. J., Spurlock, D. M., Weigel, K. A., Armentano, L. E., Connor, E. E., et al. (2018). Genome-wide association analyses based on a multiple-trait approach for modeling feed efficiency. *J. Dairy Sci.* 101, 3140–3154. doi: 10.3168/jds.2017-13364

Misztal, I. (2016). Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. *Genetics* 202, 401–409. doi: 10.1534/genetics.115.182089

Misztal, I., Legarra, A., and Aguilar, I. (2014a). Using recursion to compute the inverse of the genomic relationship matrix. *J. Dairy Sci.* 97, 3943–3952. doi: 10.3168/jds.2013-7752

Misztal, I., Lourenco, D., and Legarra, A. (2020). Current status of genomic evaluation. *J. Anim. Sci.* 98, 1–14. doi: 10.1093/jas/skaa101

Misztal, I., Tsuruta, S., Lourenco, D. A. L., Aguilar, I., Legarra, A., and Vitezica, Z. (2014b). *Manual for BLUPF90 Family of Programs*. Avaliable at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf (accessed December 1, 2020).

Pocrnic, I., Lourenco, A. L., Masuda, Y., Legarra, A., and Misztal, I. (2016). The dimensionality of genomic information and its effect on genomic prediction. *Genetics* 203, 573–581. doi: 10.1534/genetics.116.187013

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Na. Genet.* 38, 904–909. doi: 10.1038/ng1847

Risch, N., and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* 273, 1516–1517. doi: 10.1126/science.273.5281.1516

Sargolzaei, M., and Schenkel, F. S. (2009). QMSim: a large-scale genome simulator for livestock. *Bioinformatics* 25, 680–681. doi: 10.1093/bioinformatics/btp045

Sonesson, A. K., and Meuwissen, T. H. (2000). Mating schemes for optimum contribution selection with constrained rates of inbreeding. *Genet. Select. Evol.* 32, 231–248. doi: 10.1186/1297-9686-32-3-231

Stam, P. (1980). The distribution of the fraction of the genome identical by descent in finite random mating populations. *Genet. Res.* 35, 131–155. doi: 10.1017/S0016672300014002

Sul, J. H., Martin, L. S., and Eskin, E. (2018). Population structure in genetic studies: confounding factors and mixed models. *PLoS Genet.* 14:e1007309. doi: 10.1371/journal.pgen.1007309

Toosi, A., Fernando, R. L., and Dekkers, J. C. M. (2018). Genome-wide mapping of quantitative trait loci in admixed populations using mixed linear model and Bayesian multiple regression analysis. *Genet. Select. Evol.* 50, 1–13. doi: 10.1186/s12711-018-0402-1

Truong, B., Zhou, X., Shin, J., Li, J., van der Werf, J. H. J., Le, T. D., et al. (2020). Efficient polygenic risk scores for biobank scale data by exploiting phenotypes from inferred relatives. *Na. Commun.* 11:3074. doi: 10.1038/s41467-020-16829-x

Tsuruta, S., Lawlor, T. J., Lourenco, D. A. L., and Misztal, I. (2021). Bias in genomic predictions by mating practices for linear type traits in a large-scale genomic evaluation. *J. Dairy Sci.* 104, 662–677. doi: 10.3168/jds.2020-18668

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 Sullivan, P. G. (2010). International genomic evaluation methods for dairy cattle. *Genet. Select. Evol.* 42, 1–9. doi: 10.1186/1297-9686-42-7

Visscher, P. M., Brown, M. A., McCarthy, M. I., and Yang, J. (2012). Five years of GWAS discovery. *Am. J. Hum. Genet.* 90, 7–24. doi: 10.1016/j.ajhg.2011.11.02

Vitezica, Z. G., Aguilar, I., Misztal, I., and Legarra, A. (2011). Bias in genomic predictions for populations under selection. *Genet. Res.* 93, 357–366. doi: 10.1017/S001667231100022X

Wang, M., Jiang, N., Jia, T., Leach, L., Cockram, J., Waugh, L., et al. (2012). Genome-wide association mapping of agronomic and morphologic traits in highly structured populations of barley cultivars. *Theor. Appl. Genet.* 124, 233–246. doi: 10.1007/s00122-011-1697-2

Wiggans, G. R., Cooper, T. A., VanRaden, P. M., and Cole, J. B. (2011). Technical note: adjustment of traditional cow evaluations to improve accuracy of genomic predictions. *J. Dairy Sci.* 94, 6188–6193. doi: 10.3168/jds.2011-4481

Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J., et al. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* 43, 519–525. doi: 10.1038/ng.823

Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., and Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* 46, 100–106. doi: 10.1038/ng.2876

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

Conflict of Interest: 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 © 2021 Mancin, Lourenco, Bermann, Mantovani and Misztal. 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.