

RESEARCH

Open Access



The superiority of multi-trait models with genotype-by-environment interactions in a limited number of environments for genomic prediction in pigs

Hailiang Song¹, Qin Zhang² and Xiangdong Ding^{1*}

Abstract

Background: Different production systems and climates could lead to genotype-by-environment (G × E) interactions between populations, and the inclusion of G × E interactions is becoming essential in breeding decisions. The objective of this study was to investigate the performance of multi-trait models in genomic prediction in a limited number of environments with G × E interactions.

Results: In total, 2,688 and 1,384 individuals with growth and reproduction phenotypes, respectively, from two Yorkshire pig populations with similar genetic backgrounds were genotyped with the PorcineSNP80 panel. Single- and multi-trait models with genomic best linear unbiased prediction (GBLUP) and BayesC π were implemented to investigate their genomic prediction abilities with 20 replicates of five-fold cross-validation. Our results regarding between-environment genetic correlations of growth and reproductive traits (ranging from 0.618 to 0.723) indicated the existence of G × E interactions between these two Yorkshire pig populations. For single-trait models, genomic prediction with GBLUP was only 1.1% more accurate on average in the combined population than in single populations, and no significant improvements were obtained by BayesC π for most traits. In addition, single-trait models with either GBLUP or BayesC π produced greater bias for the combined population than for single populations. However, multi-trait models with GBLUP and BayesC π better accommodated G × E interactions, yielding 2.2% – 3.8% and 1.0% – 2.5% higher prediction accuracies for growth and reproductive traits, respectively, compared to those for single-trait models of single populations and the combined population. The multi-trait models also yielded lower bias and larger gains in the case of a small reference population. The smaller improvement in prediction accuracy and larger bias obtained by the single-trait models in the combined population was mainly due to the low consistency of linkage disequilibrium between the two populations, which also caused the BayesC π method to always produce the largest standard error in marker effect estimation for the combined population.

(Continued on next page)

* Correspondence: xding@cau.edu.cn

¹National Engineering Laboratory for Animal Breeding, Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Conclusions: In conclusion, our findings confirmed that directly combining populations to enlarge the reference population is not efficient in improving the accuracy of genomic prediction in the presence of $G \times E$ interactions, while multi-trait models perform better in a limited number of environments with $G \times E$ interactions.

Keywords: Combined population, Genotype-by-environment interaction, Linkage disequilibrium, Multi-trait model, Pig

Background

Genomic selection [1], which relies on linkage disequilibrium (LD) between single nucleotide polymorphisms (SNPs) and causative variants, has become a useful tool in animal breeding [2] and plant breeding [3]. Generally, the accuracy of genomic selection increases as the number of animals in the reference population increases [4]. For small reference populations, combining populations of the same breed or related breeds reportedly increases the accuracy of genomic selection in cattle and pig [2, 5–8]. However, genomic prediction using combined populations has not shown a clear advantage over genomic prediction using single populations, possibly because the LD between SNPs and causative variants is not sufficiently consistent between populations [5–7]. The failure to consider genotype-by-environment ($G \times E$) interactions between populations is another important reason. Thus, exploiting $G \times E$ interactions in combined populations could be an attractive and meaningful approach for increasing the accuracy of genomic prediction.

A $G \times E$ interaction is defined as different genotypes reacting unequally to environmental changes [9], and ignoring possible $G \times E$ interactions could lead to a reduction in genetic gains. Two models are widely used to detect $G \times E$ interactions. One is a multi-trait model, which assumes that the phenotypic expressions of a trait in various environments are different traits [10]. In this case, the genetic correlation between traits in different environments is used as the indicator of a $G \times E$ interaction [9, 11]. The other model is the reaction norm model [12], which models the trajectory of animal performance as a function of an environmental gradient. However, in a small number of environments, e.g., two or three environments, the reaction norm model captures only part of the $G \times E$ interaction due to the limited amount of environmental variation. Furthermore, it is difficult to define a suitable environmental covariate in the reaction norm model [12, 13].

With the development of genomic selection in pig breeding, more farms could be included to enlarge the reference population size in China, and joint genomic evaluation across countries or breeding organizations is expected. In dairy cattle, $G \times E$ interactions are usually ignored in combined-population genomic prediction; for example, Zhou et al. [14] reported that the consistency of LD was very high (0.97) between Chinese and Nordic

Holsteins, indicating a high level of genetic similarity between the two populations. Therefore, it is necessary to consider the influence of $G \times E$ interactions, as the consistency of LD between pig populations is relatively low compared to that in dairy cattle. The objectives of this study were to evaluate $G \times E$ interactions between two Yorkshire pig populations with similar genetic backgrounds and to investigate the performance of multi-trait models in genomic prediction in a limited number of environments with $G \times E$ interactions.

Methods

Ethics statement

The whole procedure for collecting pig blood samples was carried out in strict accordance with the protocol approved by the Animal Welfare Committee of China Agricultural University (Permit Number: DK996).

Population and phenotypes

Yorkshire pig populations were sampled from two breeding farms in the north (Beijing) and south (Fujian Province) of China, designated Beijing and Fujian, respectively, for convenience. The pigs in the Beijing and Fujian populations are American Yorkshire progeny with the same genetic background but no pedigree links. Phenotypic records of reproductive traits, namely, the number of piglets born alive (NBA) and the total number of piglets born (TNB), and growth traits, namely, days to 100 kg (AGE) and backfat thickness at 100 kg (BFT), were examined. The measurements of AGE and BFT were reported in Song et al. [7]. The phenotypic information is presented in Table 1. For the growth traits, phenotypic information from 28,827 and 13,860 pigs born in 2008–2017 and 2007–2018, respectively, in the Beijing and Fujian populations was available. For the reproductive traits, the farrow records of 5,968 and 3,115 pigs born in 2007–2017 and 2007–2018, respectively, in Beijing and Fujian were obtained.

Construction of corrected phenotypes

In pig genomic prediction analyses, corrected phenotypes derived from pedigree-based estimated breeding values (EBVs) are usually used as response variables. Conventional EBVs and genetic parameters were estimated based on a repeatability model of reproductive traits implemented separately in each population. In the

Table 1 Summary statistics for the two Yorkshire pig populations, including the numbers of genotyped animals and estimates of heritability (h^2)

| Population ^a | Trait ^b | N-obs ^c | Max | Min | SD | Genotyped animals | h^2 (SE) |
|-------------------------|--------------------|--------------------|--------|-------|-------|-------------------|-------------|
| Beijing | AGE | 28,827 | 210.98 | 124 | 13.86 | 1732 | 0.33(0.01) |
| | BFT | 28,827 | 30.74 | 5.03 | 2.43 | 1732 | 0.34 (0.01) |
| | NBA | 5,968/20,005 | 21 | 0 | 3.13 | 923 | 0.07 (0.01) |
| | TNB | 5,968/20,005 | 25 | 0 | 3.09 | 923 | 0.08 (0.01) |
| Fujian | AGE | 13,860 | 206.31 | 119.9 | 10.27 | 956 | 0.42 (0.02) |
| | BFT | 13,860 | 28.62 | 4.09 | 3.64 | 956 | 0.44 (0.02) |
| | NBA | 3,115/11,731 | 22 | 0 | 2.88 | 461 | 0.09 (0.02) |
| | TNB | 3,115/11,731 | 22 | 1 | 2.98 | 461 | 0.11 (0.02) |

^a Yorkshire pig populations from Beijing and Fujian with similar genetic backgrounds but located in the north and south of China

^b AGE: days to 100 kg; BFT: backfat thickness at 100 kg; NBA: number of piglets born alive; TNB: total number of piglets born

^c N-obs: number of individuals/observations

model, the fixed effects included herd, year and season, and the random effects included the additive genetic, random residual and permanent environmental effects. For the growth traits, a bivariate animal model was implemented, and the fixed effects included herd, year, season and sex. In addition to the additive genetic effect of each individual and the random residual effect, litter effect was also taken into account as a random effect. A total of 31,529 and 32,175 animals were traced back to construct a pedigree relationship matrix for the Beijing and Fujian populations, respectively. Afterwards, corrected phenotypic values (y_c) for reproductive traits were calculated as the EBV plus the average of estimated residuals over parities for a sow, and y_c values for growth traits were computed as the EBV plus the estimated residual for each individual in each population. EBVs and genetic parameters were computed using the DMUAI procedure in DMU software [15].

Genotype data and quality control

Genomic DNA was extracted from blood samples using a TIANamp Blood DNA Kit DP348 (Tiangen, Beijing, China). Genotyping was performed using the PorcineSNP80 BeadChip (Illumina, San Diego, CA, USA), which includes 68,528 SNPs distributed across the entire pig genome. The number of genotyped animals for each trait is presented in Table 1.

Missing genotypes for SNPs with known chromosomal positions were imputed by BEAGLE with default parameter settings [16], and those for SNPs with unknown positions were discarded. Based on the imputed dataset, SNPs were excluded from the analysis if the minor allele frequency was less than 0.05, the call frequency score was less than 0.90, or the genotype frequencies deviated from Hardy-Weinberg equilibrium with a P -value lower than 10^{-7} . The animals with a call rate less than 0.90 or with an EBV reliability less than 0.3 were removed. After quality control, 2,688 and 1,384 genotyped individuals

remained for the growth and reproductive traits, respectively, and 56,445 SNPs were ultimately used.

Principal component analysis

Principal component analysis (PCA) was performed on the **G** matrix using GCTA software [17]. This resulted in a matrix of eigenvectors in descending order that represented principal components (PCs), where PC1 had the largest eigenvalue. The overall structuring of genetic variation was visualized in a scatterplot of the top few PCs.

Measure of linkage disequilibrium

LD between a pair of SNPs was measured as r_{LD}^2 and r_{LD} [18], and r_{LD}^2 was calculated as follows:

$$r_{LD}^2 = \frac{(f(AB) - f(A)f(B))^2}{f(A)f(a)f(B)f(b)},$$

where $f(AB)$, $f(A)$, $f(a)$, $f(B)$ and $f(b)$ are the observed frequencies of haplotype AB and alleles A, a, B and b, respectively. The consistency of LD in the two populations was measured by the correlation of r_{LD} values of adjacent marker pairs on each autosome between the two populations.

Model

Four methods, namely, single-trait genomic BLUP (GBLUP), multi-trait GBLUP, single-trait BayesC π and multi-trait BayesC π , were implemented to predict the genomic EBV (GEBV) for each genotyped individual.

Single-trait GBLUP (ST-GBLUP) model

The ST-GBLUP [19] model was used to predict the GEBVs of all genotyped individuals.

$$y = \mathbf{1}\mu + \mathbf{Z}\mathbf{a} + \mathbf{e},$$

where y is the vector of corrected phenotypic values; μ

is the overall mean; $\mathbf{1}$ is a vector of 1; and \mathbf{a} is the vector of genomic breeding values, following a normal distribution $N(0, \mathbf{G} \sigma_a^2)$, in which σ_a^2 is the variance of the additive genetic effect and \mathbf{G} is the marker-based genomic relationship matrix [19]. \mathbf{e} is the vector of random errors, following a normal distribution $N(0, \mathbf{I} \sigma_e^2)$, in which σ_e^2 is the residual variance.

Multi-trait GBLUP (MT-GBLUP) model

The MT-GBLUP model was defined as.

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix},$$

where $\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$ is the vector of observed values for traits I and II (corrected phenotypic values of the same trait in different populations); \mathbf{I}_1 and \mathbf{I}_2 are the identity matrices; $\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}$ is the vector of intercepts for traits I and II; $\begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix}$ is the vector of additive genetic effects for the two traits, following a normal distribution $N(\mathbf{0}, \mathbf{G} \otimes \mathbf{M})$, where $\mathbf{M} = \begin{bmatrix} \sigma_{g1}^2 & \sigma_{g12}^2 \\ \sigma_{g12}^2 & \sigma_{g2}^2 \end{bmatrix}$ is the variance and covariance matrix of the genomic breeding values of the two traits; \mathbf{Z}_1 and \mathbf{Z}_2 are the incidence matrices associating \mathbf{g}_1 and \mathbf{g}_2 with \mathbf{y}_1 and \mathbf{y}_2 ; and $\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}$ is the vector of random errors with a distribution of $N(\mathbf{0}, \mathbf{I} \otimes \mathbf{R})$, where \mathbf{I} is the identity matrix and $\mathbf{R} = \begin{bmatrix} \sigma_{e1}^2 & \sigma_{e12}^2 \\ \sigma_{e12}^2 & \sigma_{e2}^2 \end{bmatrix}$ is the residual variance and covariance matrix of the two traits. The genetic correlation between two traits was calculated as $\frac{\sigma_{g12}^2}{\sqrt{\sigma_{g1}^2 \sigma_{g2}^2}}$.

Single-trait BayesC π (ST-BayesC π) model

In ST-BayesC π [20], marker effects on phenotypic traits were sampled from a mixture of null and normal distributions,

$$\mathbf{y}_i = \boldsymbol{\mu} + \sum_j^p m_{ij} \boldsymbol{\alpha}_j + \mathbf{e}_i$$

$$(\boldsymbol{\alpha}_j | \pi, \sigma_a^2) \begin{cases} \sim N(0, \sigma_a^2) & \text{probability}(1 - \pi) \\ 0 & \text{probability } \pi \end{cases}$$

where \mathbf{y}_i is a vector of phenotypes, $\boldsymbol{\mu}$ is a vector of overall means, m_{ij} is the genotype covariate at locus j for individual i (coded as 0, 1 and 2), p is the number of genotyped loci, $\boldsymbol{\alpha}_j$ is a vector of allele substitution effects, and \mathbf{e}_i is a vector of random residuals for individual i with a distribution $N(0, \sigma_e^2)$. The markers in the model shared a common variance σ_a^2 . The prior for the allele

substitution effects of each molecular marker $\boldsymbol{\alpha}_j$ depends on the variance σ_a^2 and the probability π that markers do not have a genetic effect.

Multi-trait BayesC π (MT-BayesC π) model

In MT-BayesC π , where each locus can have an effect on any combination of traits [21], the prior for $\boldsymbol{\alpha}_{jk}$, the allele substitution or marker effect on trait k for locus j , is a mixture with a point mass at zero and univariate normal distribution conditional on σ_k^2 :

$$\mathbf{y}_i = \boldsymbol{\mu} + \sum_j^p m_{ij} \boldsymbol{\alpha}_j + \mathbf{e}_i$$

$$(\boldsymbol{\alpha}_{jk} | \pi_k, \sigma_k^2) \begin{cases} \sim N(0, \sigma_k^2) & \text{probability}(1 - \pi_k) \\ 0 & \text{probability } \pi_k \end{cases}$$

where m_{ij} , j and p are the same as in ST-BayesC π . \mathbf{y}_i is a vector of phenotypes of k traits for individual i (corrected phenotypic values of the same trait in different populations), $\boldsymbol{\mu}$ is a vector of overall means for k traits, and $\boldsymbol{\alpha}_j$ is a vector of marker allele substitution effects on k traits for locus j . The residuals, \mathbf{e}_i , are a priori assumed to be independently and identically distributed multivariate normal vectors with a null mean and covariance matrix \mathbf{R} (\mathbf{R} is the same as described for the multi-trait GBLUP model), which, in turn, is assumed to have an inverse Wishart prior distribution, $\mathbf{W}_t^{-1}(\mathbf{S}_e, \nu_e)$. The covariance between effects for trait k and k' at the same locus, i.e., $\boldsymbol{\alpha}_{jk}$ and $\boldsymbol{\alpha}_{jk'}$ is

$$\text{cov}(\boldsymbol{\alpha}_{jk}, \boldsymbol{\alpha}_{jk'} | \sigma_{kk'}) \begin{cases} \sigma_{kk'} & \text{if both } \boldsymbol{\alpha}_{jk} \neq 0 \text{ and } \boldsymbol{\alpha}_{jk'} \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

Imputation of missing phenotypic data was implemented in each Markov chain Monte Carlo (MCMC) iteration in MT-BayesC π .

For the ST-BayesC π and MT-BayesC π methods, the MCMC chain was run for 20,000 cycles, and the first 10,000 cycles were discarded as burn-in. Every 10th sample of the remaining 10,000 iterations was saved to estimate SNP effects and the variance components. The Julia package JWAS was used for BayesC π analyses [22], and the DMUAI procedure, implemented in DMU software, was used for GBLUP analyses.

Cross-validation and prediction accuracy

In this study, the accuracy and unbiasedness of prediction were obtained through 5-fold cross-validation (CV) for the Beijing and Fujian populations, where the Beijing (Fujian) population was randomly split into 5 folds, predicting one fold based on the other 4 folds, the combined population was divided into 4 folds and another population, or a multi-trait model was used in which the values of the same trait in 4

folds and another population were considered different traits. In each round of CV, phenotypes from one fold (validation population) were removed from the dataset, and the remaining folds (reference population) were used to predict the future performance of animals in the validation population. This 5-fold CV was replicated 20 times, and the results are presented as the mean and standard deviation for the 20 replicates. Here, the single-trait models (ST-GBLUP and ST-BayesC π) based on single populations and the combined population were termed ST-GBLUP(BayesC π)_single and ST-GBLUP(BayesC π)_combined, respectively. For ST-GBLUP(BayesC π)_single, GBLUP(-BayesC π)_combined and MT-GBLUP(BayesC π), the validation population was the same for the four approaches in each replicate of 5-fold CV. The accuracy of genomic prediction was evaluated as $r(y_c, \text{GEBV})$, the correlation between GEBVs and corrected phenotypic values y_c in the validation population. In addition, $b(y_c, \text{GEBV})$, the regression of y_c on GEBVs, was also calculated to assess the possible inflation or unbiasedness of predictions. A two-tailed paired t-test was used to compare prediction accuracy between a pair of scenarios.

Results

Population structure and genetic parameters

To assess the population structure of the two Yorkshire pig populations, PCA was performed. Figure 1 illustrates that the genetic backgrounds of the Beijing and Fujian populations were similar, and both

populations were composed of American Yorkshire progeny. In addition, the Yorkshire population (874 pigs) with a British origin was selected for PCA, which also showed that Beijing and Fujian had similar genetic backgrounds (Additional file: Fig. S1). However, no pedigree connection between these populations was detected according to their pedigree records due to a lack of genetic exchange.

LD between a pair of SNPs was measured as r_{LD}^2 and r_{LD} in the Beijing and Fujian populations. The LD between adjacent markers in the Beijing and Fujian populations was also investigated, as shown in Table 2. The mean r_{LD}^2 of adjacent SNP pairs within a chromosome ranged from 0.52 to 0.63 in the Beijing population and from 0.51 to 0.63 in the Fujian population. The mean r_{LD}^2 across all chromosomes was 0.56 in both the Beijing and Fujian populations, also indicating the similar genetic backgrounds of the populations. However, the consistency of LD between the two populations was not high. The correlation of r_{LD} for adjacent SNPs between the two populations ranged from 0.47 to 0.60 across all chromosomes, with a mean of 0.55 at an average marker distance of 41 Kb, indicating the insufficient consistency in LD between these two pig populations.

Estimates of heritability for the growth and reproductive traits in the two Yorkshire pig populations are shown in Table 1. In the Beijing population, the heritabilities of AGE and BFT were 0.33 and 0.34, respectively, with a standard error of 0.01. The heritabilities of AGE and BFT were 0.42 and 0.44, respectively, with a standard error of 0.02 in the Fujian population. For NBA and TNB, the

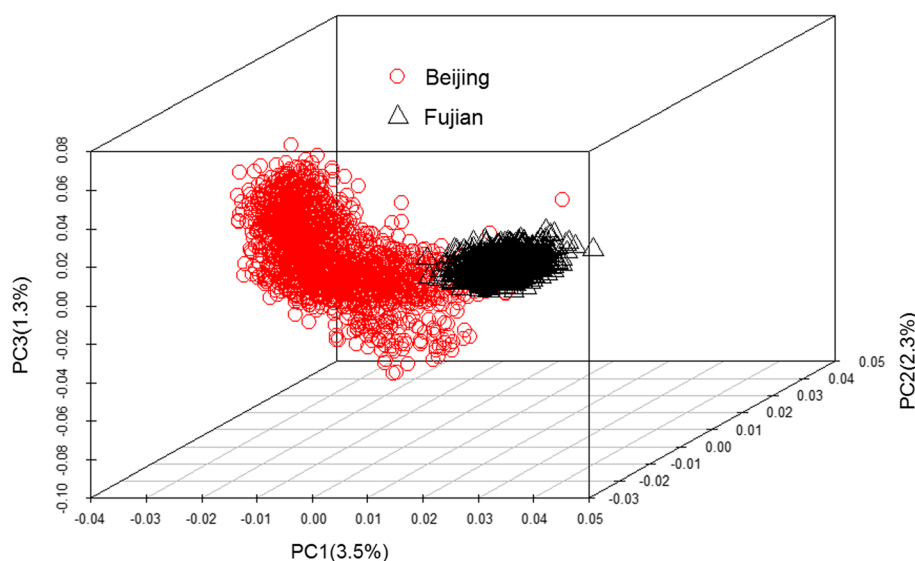


Fig. 1 Principal component analysis (PCA) of two Yorkshire pig populations. Beijing and Fujian represent two Yorkshire pig populations from two Chinese pig breeding farms; PC1 (3.5%) = first principal component (variance explained by PC1 = 3.5%); PC2 (2.3%) = second principal component (variance explained by PC2 = 2.3%); PC3 (1.3%) = third principal component (variance explained by PC3 = 1.3%)

Table 2 Distance and LD (r_{LD}^2) of adjacent SNPs for each autosome (CHR)

| CHR | Length, Mb | Number of SNPs | Mean distance, kb | Mean r_{LD}^2 | | Cor ^a |
|------|-----------------------|---------------------|-------------------|----------------------|---------------------|------------------|
| | | | | Beijing ^c | Fujian ^c | |
| 1 | 300.49 | 5,525 | 55.70 | 0.63 | 0.63 | 0.55 |
| 2 | 155.0 | 3,750 | 42.34 | 0.56 | 0.54 | 0.51 |
| 3 | 138.07 | 3,267 | 43.29 | 0.54 | 0.53 | 0.54 |
| 4 | 136.82 | 3,453 | 40.59 | 0.57 | 0.57 | 0.51 |
| 5 | 106.34 | 2,671 | 40.78 | 0.54 | 0.54 | 0.60 |
| 6 | 150.43 | 3,967 | 38.84 | 0.61 | 0.60 | 0.47 |
| 7 | 128.52 | 3,563 | 36.95 | 0.54 | 0.53 | 0.58 |
| 8 | 141.61 | 3,191 | 45.46 | 0.57 | 0.56 | 0.56 |
| 9 | 146.54 | 3,533 | 42.48 | 0.55 | 0.57 | 0.58 |
| 10 | 75.40 | 2,522 | 30.63 | 0.55 | 0.53 | 0.51 |
| 11 | 83.61 | 2,105 | 40.69 | 0.52 | 0.51 | 0.50 |
| 12 | 60.64 | 2,247 | 27.65 | 0.57 | 0.56 | 0.60 |
| 13 | 208.48 | 4,026 | 53.04 | 0.59 | 0.60 | 0.58 |
| 14 | 146.71 | 3,756 | 40.01 | 0.59 | 0.59 | 0.57 |
| 15 | 150.35 | 3,208 | 48.01 | 0.58 | 0.59 | 0.52 |
| 16 | 82.85 | 2,154 | 39.41 | 0.53 | 0.52 | 0.51 |
| 17 | 66.13 | 1,938 | 34.96 | 0.58 | 0.58 | 0.53 |
| 18 | 58.37 | 1,569 | 38.12 | 0.54 | 0.52 | 0.55 |
| Mean | 2,336.36 ^b | 56,445 ^b | 41.05 | 0.56 | 0.56 | 0.55 |

^a Cor: the correlation of r_{LD} values of adjacent SNP pairs between two populations;

^b Sum over 18 autosomes

^c Yorkshire pig populations from Beijing and Fujian with similar genetic backgrounds

estimated heritabilities were 0.07 and 0.08 with a standard error of 0.01 in the Beijing population and 0.09 and 0.11 with a standard error of 0.02 in the Fujian population.

G × E interactions

Figure 2 shows the estimated genetic correlations between the Beijing and Fujian populations based on the MT-GBLUP model using all genotyped animals. For AGE and BFT, the estimates of genetic correlations were 0.618 and 0.623, respectively, with standard errors of 0.145 and 0.134. The genetic correlations of NBA and TNB were 0.714 and 0.723, respectively, with standard errors of 0.153 and 0.159. These genetic correlations indicated that G × E interactions most likely exist between the Beijing and Fujian populations, as Robertson [23] suggested that 0.80 was the threshold of biological importance for G × E interactions.

Tables 3 and 4 and Fig. 3 demonstrate the accuracies of genomic prediction for growth and reproductive traits achieved by applying single-trait and multi-trait models with GBLUP and BayesC π . Generally, in the same

scenario, the prediction accuracies of each method were the largest for BFT, as it had the highest estimated heritability among the four traits. Lower prediction accuracies were obtained for two reproductive traits, NBA and TNB, due to their low heritabilities. For the two pig populations, higher accuracies for growth traits were acquired for the Beijing population, as it had a much larger reference population than the Fujian population (Table 1), while their small reference populations for reproductive traits resulted in comparably low accuracies of genomic prediction. Tables 3 and 4 and Fig. 3 further show the superiority of the multi-trait model for genomic prediction when dealing with G × E interactions, as discussed below.

Comparison of the combined population with the single populations

Table 3 presents the accuracy of genomic prediction for growth and reproductive traits achieved by applying the GBLUP method, as measured by 20 replicates of five-fold CV. For the Beijing population, the prediction accuracies obtained with ST-GBLUP_single were 0.315 and 0.331 for AGE and BFT and 0.142 and 0.172 for NBA and TNB, respectively. When ST-GBLUP_combined was used, the genomic prediction accuracies were improved by 1.1% and 1.7% for NBA and TNB, respectively, compared with those obtained with ST-GBLUP_single, while there were no significant improvements for the growth traits AGE and BFT. In contrast, for the Fujian population, the genomic prediction accuracies for both growth and reproductive traits increased when the single populations were combined to enlarge the reference population. On average, ST-GBLUP_combined yielded 1.1% and 2.0% higher accuracies than ST-GBLUP_single for the growth and reproductive traits, respectively. Obviously, the gains in accuracy obtained for the Fujian population were larger.

However, the single-trait model with the BayesC π method did not yield higher prediction accuracies for the combined population compared with the single populations for most traits, even though the reference population was larger. As shown in Table 4, when predicting the traits of pigs from the Beijing population, the prediction accuracies obtained with ST-BayesC π _single were 0.306, 0.328, 0.156 and 0.179 for AGE, BFT, NBA and TNB, respectively. ST-BayesC π _combined produced almost the same accuracies as ST-BayesC π _single in all scenarios, and the accuracy obtained by ST-BayesC π _combined was even slightly lower in some cases; e.g., the prediction accuracies of ST-BayesC π _single and ST-BayesC π _combined for NBA were 0.156 and 0.154, respectively. A trend of similar accuracies for ST-BayesC π between the combined and single

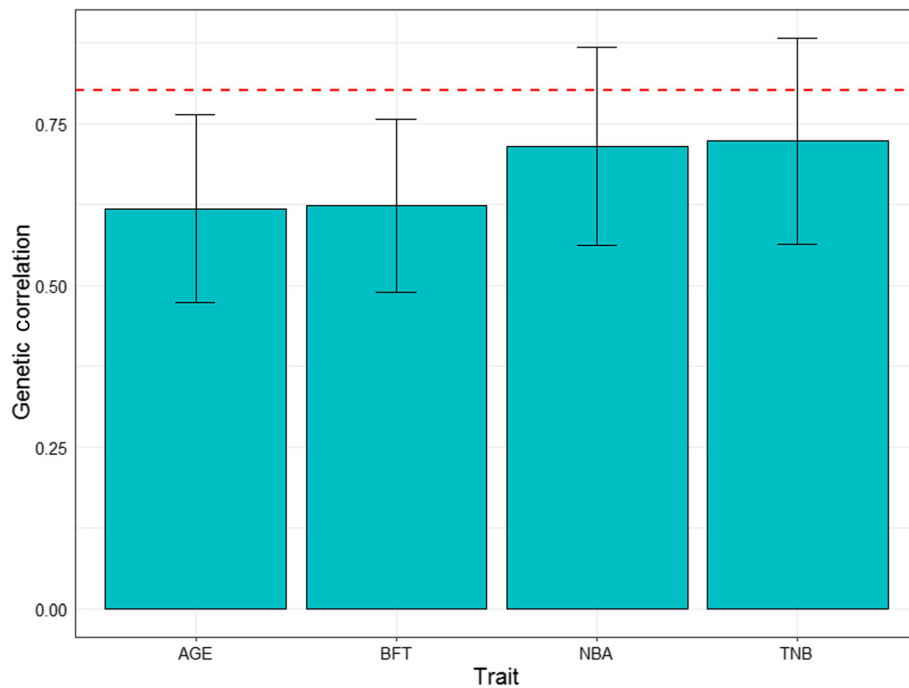


Fig. 2 Genetic correlations between the two Yorkshire pig populations obtained using the multi-trait GBLUP method with all genotyped animals. AGE: days to 100 kg; BFT: backfat thickness at 100 kg; NBA: number of piglets born alive; TNB: total number of piglets born. The red line represents the threshold of 0.8. The error bar represents the standard error

Table 3 Accuracy and unbiasedness of genomic prediction of growth and reproductive traits performed with the GBLUP method as assessed with 20 replicates of five-fold CV

| Population ¹ | Method ² | Measurement ³ | AGE ⁴ | BFT ⁴ | NBA ⁴ | TNB ⁴ |
|-------------------------|---------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| Beijing | ST-GBLUP_single | Accuracy | 0.315 ± 0.035 ^a | 0.331 ± 0.026 ^a | 0.142 ± 0.045 ^a | 0.172 ± 0.043 ^a |
| | | Unbiasedness | 1.001 ± 0.203 | 1.126 ± 0.254 | 1.220 ± 0.413 | 1.146 ± 0.411 |
| | ST-GBLUP_combined | Accuracy | 0.315 ± 0.037 ^a | 0.334 ± 0.029 ^a | 0.153 ± 0.052 ^b | 0.189 ± 0.052 ^b |
| | | Unbiasedness | 1.119 ± 0.222 | 0.827 ± 0.249 | 0.698 ± 0.360 | 0.799 ± 0.284 |
| Fujian | MT-GBLUP | Accuracy | 0.326 ± 0.032 ^b | 0.346 ± 0.025 ^b | 0.167 ± 0.043 ^c | 0.191 ± 0.044 ^b |
| | | Unbiasedness | 0.994 ± 0.196 | 1.083 ± 0.225 | 1.164 ± 0.362 | 1.112 ± 0.278 |
| | ST-GBLUP_single | Accuracy | 0.245 ± 0.017 ^a | 0.261 ± 0.035 ^a | 0.159 ± 0.053 ^a | 0.146 ± 0.052 ^a |
| | | Unbiasedness | 1.047 ± 0.267 | 1.046 ± 0.263 | 1.047 ± 0.442 | 1.050 ± 0.421 |
| ST-GBLUP_combined | Accuracy | 0.255 ± 0.023 ^b | 0.272 ± 0.043 ^b | 0.171 ± 0.062 ^b | 0.173 ± 0.055 ^b | |
| | Unbiasedness | 1.352 ± 0.303 | 1.274 ± 0.271 | 1.352 ± 0.454 | 1.191 ± 0.465 | |
| MT-GBLUP | Accuracy | 0.273 ± 0.019 ^c | 0.297 ± 0.031 ^c | 0.181 ± 0.054 ^c | 0.184 ± 0.057 ^c | |
| | Unbiasedness | 1.024 ± 0.243 | 1.013 ± 0.232 | 1.032 ± 0.433 | 1.046 ± 0.397 | |

¹ Yorkshire pig populations from Beijing and Fujian with similar genetic backgrounds

² ST-GBLUP_single: single-trait GBLUP model with a single population as the reference population; ST-GBLUP_combined: single-trait GBLUP model two populations combined as the reference population; MT-GBLUP: multi-trait GBLUP model in which values of the same trait in different populations were considered different traits

³ Accuracy: the correlation between GEBV and corrected phenotypic values in the validation population; Unbiasedness: the regression of corrected phenotypic values on GEBVs

⁴ AGE: days to 100 kg; BFT: backfat thickness at 100 kg; NBA: number of piglets born alive; TNB: total number of piglets born

^{a, b, c} Values with different superscript letters significantly differ ($P < 0.05$)

Table 4 Accuracy and unbiasedness of genomic prediction of growth and reproductive traits performed with the BayesC π method as assessed with 20 replicates of five-fold CV

| Population ¹ | Method ² | Measurement ³ | AGE ⁴ | BFT ⁴ | NBA ⁴ | TNB ⁴ |
|-------------------------|---------------------------|--------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| Beijing | ST-BayesC π _single | Accuracy | 0.306 ± 0.028 ^a | 0.328 ± 0.023 ^a | 0.156 ± 0.032 ^a | 0.179 ± 0.035 ^a |
| | | Unbiasedness | 1.218 ± 0.278 | 1.013 ± 0.212 | 1.182 ± 0.343 | 1.112 ± 0.302 |
| | ST-BayesC π _combined | Accuracy | 0.304 ± 0.032 ^a | 0.328 ± 0.038 ^a | 0.154 ± 0.035 ^a | 0.180 ± 0.034 ^a |
| | | Unbiasedness | 1.287 ± 0.265 | 0.853 ± 0.274 | 1.152 ± 0.362 | 1.124 ± 0.293 |
| | MT-BayesC π | Accuracy | 0.312 ± 0.035 ^b | 0.347 ± 0.026 ^b | 0.183 ± 0.029 ^b | 0.201 ± 0.031 ^b |
| | | Unbiasedness | 1.160 ± 0.217 | 1.056 ± 0.224 | 1.065 ± 0.236 | 0.992 ± 0.272 |
| Fujian | ST-BayesC π _single | Accuracy | 0.243 ± 0.023 ^a | 0.262 ± 0.024 ^a | 0.164 ± 0.043 ^a | 0.152 ± 0.044 ^a |
| | | Unbiasedness | 1.052 ± 0.263 | 1.032 ± 0.232 | 1.115 ± 0.321 | 1.076 ± 0.362 |
| | ST-BayesC π _combined | Accuracy | 0.244 ± 0.025 ^a | 0.264 ± 0.028 ^a | 0.165 ± 0.047 ^a | 0.154 ± 0.047 ^a |
| | | Unbiasedness | 0.874 ± 0.134 | 1.330 ± 0.254 | 1.245 ± 0.308 | 1.222 ± 0.353 |
| | MT-BayesC π | Accuracy | 0.275 ± 0.024 ^b | 0.284 ± 0.021 ^b | 0.181 ± 0.051 ^b | 0.183 ± 0.049 ^b |
| | | Unbiasedness | 1.034 ± 0.197 | 1.088 ± 0.256 | 1.143 ± 0.363 | 1.059 ± 0.345 |

¹ Yorkshire pig populations from Beijing and Fujian with similar genetic backgrounds

² ST-GBLUP_single: single-trait GBLUP model with a single population as the reference population; ST-GBLUP_combined: single-trait GBLUP model two populations combined as the reference population; MT-GBLUP: multi-trait GBLUP model in which values of the same trait in different populations were considered different traits

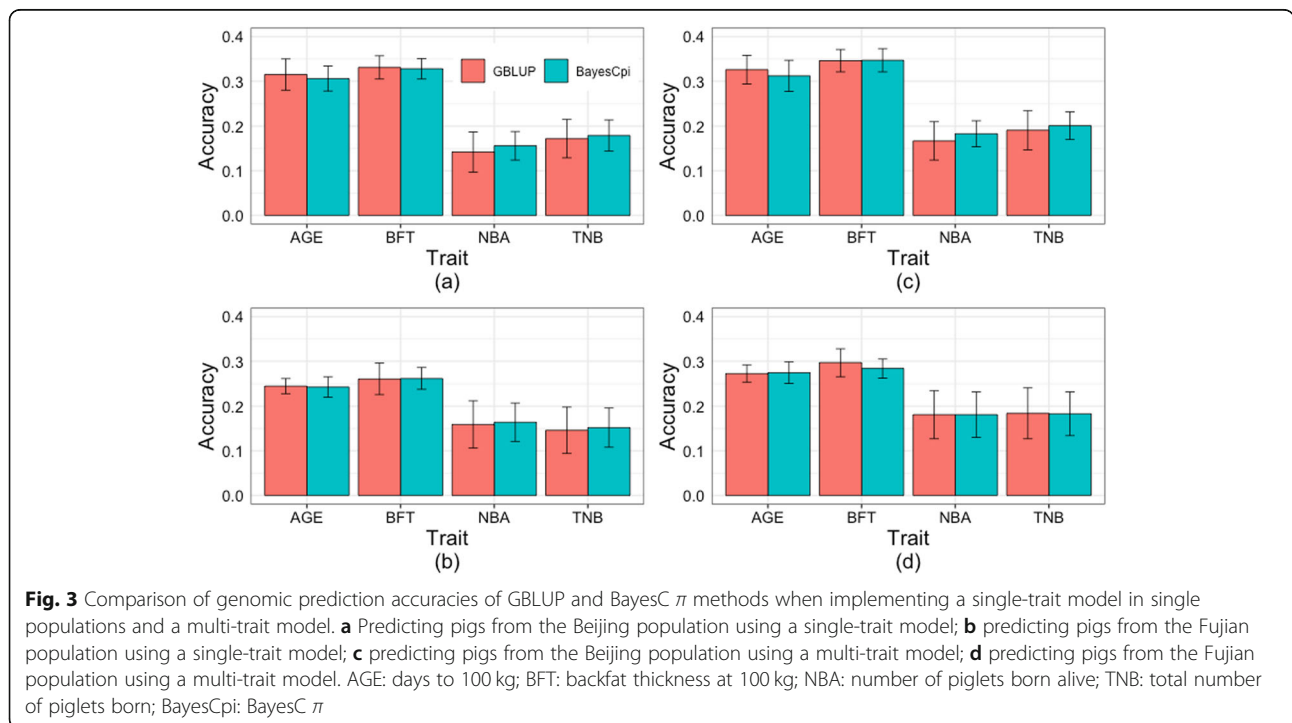
³ Accuracy: the correlation between GEBVs and corrected phenotypic values in the validation population; Unbiasedness: the regression of corrected phenotypic values on GEBVs

⁴ AGE: days to 100 kg; BFT: backfat thickness at 100 kg; NBA: number of piglets born alive; TNB: total number of piglets born

^{a, b} Values with different superscript letters significantly differ ($P < 0.05$)

populations was also found in the Fujian population, as demonstrated in Table 4. The average prediction accuracies obtained with ST-BayesC π _single and ST-BayesC π _combined were almost the same for the growth traits and reproductive traits.

The unbiasedness of genomic prediction of growth and reproductive traits as assessed with 20 replicates of five-fold CV is shown in Tables 3 and 4. When predicting the traits of pigs from the Beijing population based on the GBLUP method, larger bias (0.199) was produced



by ST-GBLUP_combined than by ST-GBLUP_single. Likewise, unbiasedness was close to 1 with the ST-GBLUP_single method for growth and reproductive traits when predicting the traits of pigs from the Fujian population, while ST-GBLUP_combined yielded a 0.292 larger bias than ST-GBLUP_single (Table 3). Similarly, ST-BayesC π also produced more bias for the combined population than for the single populations in most situations (Table 4).

In addition, prediction accuracies obtained with training on the Beijing population and validation in the Fujian population and training on the Fujian population and validation in the Beijing population using ST-GBLUP and ST-BayesC π were determined. Low prediction accuracies and large bias were observed in all scenarios (Additional file: Table S1), indicating that using one population to predict another is infeasible, even though the genetic backgrounds of the Beijing and Fujian populations were similar in this study.

Comparison of the multi-trait model with the single-trait model

Both single-trait and multi-trait models with GBLUP and BayesC π were implemented to address the combined population in this study. Tables 3 and 4 show the accuracies and unbiasedness of genomic prediction obtained with the multi-trait and single-trait models in the same scenarios. In general, the multi-trait model showed significant superiority over the single-trait model. For the GBLUP method, as shown in Table 3, when predicting the traits of pigs from the Beijing population, the multi-trait GBLUP model (MT-GBLUP) yielded approximately 1.3% (1.2%) and 2.2% (0.8%) higher accuracies than ST-GBLUP_single (ST-GBLUP_combined) for the growth and reproductive traits, respectively. Furthermore, a small amount of bias was obtained with MT-GBLUP in most scenarios. A similar trend was found in the Fujian population. As shown in Table 3, MT-GBLUP yielded 3.2% (2.2%) and 3.0% (1.1%) higher accuracies on average for the growth and reproductive traits, respectively, than ST-GBLUP_single (ST-GBLUP_combined). The unbiasedness was close to 1 when using the MT-GBLUP methods for growth and reproductive traits in most situations. Again, the gains obtained by MT-GBLUP in the Fujian population were larger than those in the Beijing population.

For the multi-trait model with BayesC π , as shown in Table 4, when predicting the traits of pigs from the Beijing population, MT-BayesC π yielded approximately 1.3% (1.4%) and 2.5% (2.5%) higher accuracies than ST-BayesC π _single (ST-BayesC π _combined) for the growth and reproductive traits, respectively. The largest gain over ST-BayesC π _single and ST-BayesC π _combined was 2.7% and 2.9% for NBA, respectively. When

predicting the traits of pigs from the Fujian population, MT-BayesC π obtained higher accuracies for all traits (Table 4), yielding 2.7% (2.6%) and 2.4% (2.3%) gains on average over ST-BayesC π _single (ST-BayesC π _combined) for the growth and reproductive traits, respectively. MT-BayesC π also produced bias as low as that of ST-BayesC π _single in most scenarios.

Comparison of the GBLUP and BayesC π methods

Figure 3 presents the accuracy of genomic prediction of growth and reproductive traits achieved by applying single-population ST-GBLUP (BayesC π) and MT-GBLUP (BayesC π) methods. When predicting the traits of pigs from the Beijing population, GBLUP yielded approximately 1.0% higher accuracy than BayesC π regardless of whether the single- or multi-trait model was used for AGE and a bias close to 0 with the GBLUP method in this scenario, while there was no difference in accuracy between GBLUP and BayesC π for BFT. For reproductive traits in the Beijing population, the accuracy of genomic prediction was increased by 1.3% on average for MT-BayesC π compared to MT-GBLUP, and no difference in accuracy was found between ST-GBLUP and ST-BayesC π . When predicting the traits of pigs from the Fujian population, GBLUP and BayesC π produced similar prediction accuracies for AGE; e.g., the prediction accuracies with ST-GBLUP_single and ST-BayesC π _single were 0.245 and 0.243, respectively, while less bias was obtained by GBLUP in this scenario. There was no difference in prediction accuracy or unbiasedness between ST-GBLUP_single and ST-BayesC π _single for BFT, while MT-GBLUP achieved a 1.3% higher accuracy than MT-BayesC π . For reproductive traits in the Fujian population, there was no difference in prediction accuracy between GBLUP and BayesC π , except that MT-GBLUP_combined achieved a 1.9% higher accuracy than MT-BayesC π _combined for TNB. Generally, GBLUP performed comparably to BayesC π in the single-trait and multi-trait models.

Discussion

In this study, we investigated the accuracy of genomic prediction in two Yorkshire pig populations with similar genetic backgrounds when $G \times E$ interactions were taken into account. Our results revealed $G \times E$ interactions between the two populations for the reference growth and reproductive traits. We further explored whether directly combining populations to enlarge the reference population is efficient in improving the accuracy of genomic prediction in the presence of $G \times E$ interactions. Our results demonstrated the superiority of the multi-trait model for genomic prediction in dealing with $G \times E$ interactions, which could better accommodate the $G \times E$ interactions. Higher prediction accuracies and lower bias

were obtained with the multi-trait model than with the single-trait model when implementing either GBLUP or BayesC π in a given situation.

$G \times E$ interactions play an important role in pig populations, and they should be considered in breeding programs to select the best animals in different environments [11, 24]. The detection of $G \times E$ interactions relies on a genetic correlation of 0.8 for one trait in different environments, the threshold suggested by Robertson [23]. Accordingly, $G \times E$ interactions between two pig populations for growth and reproductive traits (ranging from 0.618 to 0.723) were observed. Similarly, Li et al. [25] reported on the across-country genetic correlations (ranging from 0.604 to 0.726) of dairy cattle, indicating that an important $G \times E$ interaction exists between Brazilian and Nordic (or Nordic and French) populations. Hay and Roberts [26] also reported the existence of genotype-by-nutritional environment interactions for growth and carcass traits in beef cattle with genetic correlations below the threshold value of 0.8. In this study, SNP markers were used to construct the relationships (G matrix) between the two populations in order to estimate the genetic correlations, as no linked pedigree was available. Compared with a pedigree used to estimate genetic correlations, the realized relationships among individuals are captured by marker information and achieve more accurate estimates of genetic correlations. However, the number of genotyped animals was not large enough to produce a lower standard error for genetic correlations compared to that obtained by estimating heritability. We therefore used SNP markers to assess the population structure of the two pig populations, and the similar patterns of LD on each chromosome between these two populations further showed that they had similar genetic backgrounds. However, the correlation of r_{LD} values of adjacent SNPs with a mean of 0.55 indicated the insufficient consistency in LD between these two pig populations, which could be due to $G \times E$ interactions. Genetic correlations are caused by the pleiotropic action of genes and linkage between genes affecting different traits [9]. If the performance values of the same trait in different populations (environments) are regarded as different traits and are affected by different genes (that are of course pleiotropic), the phase of linkage of these genes and the consistency of their linkage (not LD) in the two populations will affect the strength of the genetic correlation and $G \times E$ interaction.

The size of the reference population is one key factor in genomic prediction [4]. Combining populations is an easy but practical way to improve the accuracy of genomic prediction, and its advantages in practice have been reported by EuroGenomics [8] and North American [2] consortiums. Similar results were shown in this

study, in which higher prediction accuracies were obtained by the single-trait model for the combined population than for the single populations in most situations (Table 3), and such gains were greater for the population with a small reference size; e.g., the prediction accuracy in the combined population was 1.1% higher than that in the single population for the growth trait BFT in the Fujian population but only 0.3% higher in the Beijing population, and the number of genotyped animals in the Beijing population was much larger than that in the Fujian population (Table 2), in accordance with the features of other studies [5, 14]. However, in some situations, accuracy was not increased or slightly decreased in the combined population; e.g., the prediction accuracies for AGE in the Beijing population obtained with ST-BayesC π _combined and ST-BayesC π _single were 0.306 and 0.304, respectively (Table 4). Generally, in this study, the single-trait model exhibited only a slight gain in genomic prediction accuracy for most traits when using the combined population. In addition, the single-trait model produced more bias for the combined population than for the single populations (Tables 3 and 4). The lower improvement or slight decrease in the combined population compared with the expectation could be due to the weak consistency in LD between populations, as pointed out by Lund et al. [8].

Many studies have shown that inconsistent LD between SNPs and causative variants across populations causes disadvantages compared with using a combined population in genomic prediction [5, 6, 27, 28]. In this study, the correlation of r_{LD} values of adjacent SNPs with a mean of 0.55 indicated the insufficient consistency in LD between the Beijing and Fujian populations. Weak consistency in LD will lead to bias in SNP effect estimates when populations are simply combined with a single-trait model, further resulting in fewer improvements in accuracy and larger bias. Therefore, the effect of increasing the consistency in LD or reducing its noise on GEBVs is worth investigating. On the one hand, the consistency in LD between populations could be affected by the density of chip SNPs. De Roos et al. [29] found that markers in LD with causative variants across diverged breeds would require $\sim 300,000$ SNPs. Hoze et al. [30] reported that 2% higher prediction accuracies were acquired for combined small populations of dairy cattle breeds when using the 777 K panel compared to the 54 K panel, and our previous study showed that imputed whole-genome sequencing data hold the potential to increase the accuracy of genomic prediction for combined populations in pigs [7].

On the other hand, reasonable models should be explored to take advantage of the correlation of LD between populations. The single-trait model ignores the inconsistency in LD between populations, while the

multi-trait model treats populations as different environments and uses covariance to consider the correlation between them, which is caused by the consistency (no matter how high or low) in LD between populations. In this study, the multi-trait model performed best in all scenarios (Tables 3 and 4). Similar results were obtained in a Brazilian Holstein population by adding data from Nordic and French Holstein populations and using a multi-trait model [25]. In addition, our results also showed that less bias was obtained with the multi-trait model compared to the single-trait model, indicating that the multi-trait model can not only use the combined information of populations to increase prediction accuracy but also reduce the impact of LD inconsistencies.

The multi-trait model and reaction norm model are two prevalent methods for handling $G \times E$ interactions in genetic evaluations. Only a few studies have explicated the advantages and disadvantages of these two kinds of methods. Falconer et al. [9] reported that $G \times E$ interactions could be detected using a multi-trait model, in which the trait values in each environment are treated as genetically distinct traits. In a limited number of environments (e.g., in this study, with two breeding farms), a multi-trait model could capture the $G \times E$ interaction between these environments. However, the computational demand of multi-trait models will increase rapidly with an increase in the number of environmental levels, as more (co)variance components will have to be estimated and convergence will become increasingly difficult. In contrast to the multi-trait model, the reaction norm model can handle a large number of or continuous environmental levels (e.g., farm, year and seasonal effects), and genetic parameters could be estimated at each environmental level [12]. However, the reaction norm model captures only part of the $G \times E$ interaction due to the limited number of environmental levels. In this study, the genomic prediction accuracies were not improved by using the reaction norm model with farm, year and seasonal effects as environmental covariates due to the limited number of environments (results not shown). Therefore, the reaction norm model could not perform well in a limited number of environments. In addition, convergence of the reaction norm model could

be hindered because of the complexity of the model when the data are less informative [11]. Therefore, a multi-trait model is optimal for analyzing $G \times E$ interactions in a limited number of environments. Furthermore, the multi-trait model is more convenient in practical breeding, and it can accommodate phenotypes that are not measured in exactly the same way, e.g., different scales of deregressed proofs and the existence of $G \times E$ interactions in Chinese and Nordic Holsteins [25, 31]. Such phenotypes cannot be analyzed by a single-trait model for joint genomic prediction, while the use of a multi-trait model is plausible in such cases.

In this study, GBLUP and BayesC π , assuming a common (co)variance for all (GBLUP) or a proportion (BayesC π) of the SNP effects, were used in this study. As many studies have reported for real data [32, 33], the average prediction accuracies were not significantly different between GBLUP and BayesC π in the single populations and multi-trait model in this study (Fig. 3). However, BayesC π performed worse than GBLUP for the combined population, and no improvements in accuracy were obtained by BayesC π for most traits compared to that obtained with a single population. In addition, BayesC π produced a larger bias (Table 4). Table 5 provides the standard error of the marker effect estimated by BayesC π in single and combined populations. The BayesC π method for the combined population produced the largest standard error of the marker effect in all situations, implying larger bias when estimating genomic breeding values. A larger standard error for the marker effect might have been caused by the inconsistency in LD between the Beijing and Fujian populations. In contrast, the minimum standard error of the marker effects estimated by the multi-trait BayesC π model also indicated the superiority of multi-trait BayesC π in genomic prediction. Multi-trait BayesC π was first proposed by Jia and Jannink [34], with the restrictive assumption that a locus simultaneously affects all the traits or none of them. In this study, multi-trait BayesC π allowed a locus to affect any combination of traits, which is biologically realistic, especially in multi-trait analyses involving many traits [21]. However, it might not be possible to apply multi-trait single-step GBLUP (BayesC π) [35–

Table 5 Mean value of the standard error of marker effects estimated by the BayesC π method using all genotyped animals

| Trait ^b | Single-trait model | | | Multi-trait model |
|--------------------|---------------------------------|--------------------------------|---------------------|-------------------|
| | Beijing ^a population | Fujian ^a population | Combined population | |
| AGE | 0.0062 | 0.0079 | 0.0093 | 0.0053 |
| BFT | 0.0050 | 0.0069 | 0.0096 | 0.0048 |
| NBA | 0.0039 | 0.0047 | 0.0050 | 0.0039 |
| TNB | 0.0042 | 0.0051 | 0.0054 | 0.0038 |

^a Yorkshire pig populations from Beijing and Fujian with similar genetic backgrounds

^b AGE: days to 100 kg; BFT: backfat thickness at 100 kg; NBA: number of piglets born alive; TNB: total number of piglets born

37] in this study, as no pedigree link between the two pig populations was available and the BayesC π method was too computationally demanding.

Conclusions

$G \times E$ interactions constitute a potential source of inefficiency in animal breeding. Ignoring possible $G \times E$ interactions could lead to a reduction in genetic gains. Our results demonstrated that directly combining populations to enlarge the reference population is not efficient in improving the accuracy of genomic prediction in the presence of $G \times E$ interactions in two Yorkshire pig populations, while the multi-trait model performed better in a limited number of environments with weak $G \times E$ interactions. This study will be helpful when leveraging $G \times E$ interactions in practical genomic selection.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s40104-020-00493-8>.

Additional file 1: Fig. S1. Principal component analysis (PCA) of the Beijing and Fujian Yorkshire pig populations and a British Yorkshire pig population.

Additional file 2: Table S1. Accuracy and unbiasedness of genomic prediction of growth and reproductive traits obtained when using one population to predict another.

Abbreviations

$G \times E$: Genotype-by-environment; GBLUP: Genomic best linear unbiased prediction; LD: Linkage disequilibrium; SNPs: Single nucleotide polymorphisms; Beijing and Fujian represent two Yorkshire pig populations from two Chinese pig breeding farms; AGE: Days to 100 kg; BFT: Backfat thickness at 100 kg; NBA: Number of piglets born alive; TNB: Total number of piglets born; y_c : Corrected phenotypic values; GEBV: Genomic EBV; ST-GBLUP(BayesC π): Single-trait GBLUP(BayesC π); MT-GBLUP(BayesC π): Multi-trait GBLUP(BayesC π); CV: Cross validation; ST-GBLUP(BayesC π)_single: ST-GBLUP(BayesC π) based on single population; ST-GBLUP(BayesC π)_combined: ST-GBLUP(BayesC π) based on combined population; PCA: Principal component analysis; PC: Principal component; MCMC: Markov chain Monte Carlo

Acknowledgements

The authors are grateful to Beijing Liuma and Fujian Yongcheng Pig Breeding Co., Ltd. for providing blood samples.

Authors' contributions

HLS and XDD designed the experiments. HLS performed statistical analysis and wrote the manuscript. QZ and XDD revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the earmarked fund for China Agriculture Research System (CARS-35), Modern Agriculture Science and Technology Key Project of Hebei Province (19226376D), the National Key Research and Development Project (SQ2019YFE00771), the National Natural Science Foundation of China (31671327). Major Project of Selection for New Livestock and Poultry Breeds of Zhejiang Province (2016C02054-5).

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The experiments involving animals were approved by the Animal Welfare Committee of China Agricultural University. There was no use of human participants, data or tissues.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹National Engineering Laboratory for Animal Breeding, Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China. ²Shandong Provincial Key Laboratory of Animal Biotechnology and Disease Control and Prevention, Shandong Agricultural University, Taian 271001, China.

Received: 1 March 2020 Accepted: 7 July 2020

Published online: 19 August 2020

References

1. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157(4):1819–29.
2. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, et al. Invited review: reliability of genomic predictions for north American Holstein bulls. *J Dairy Sci*. 2009;92(1):16–24.
3. Zhong S, Dekkers JC, Fernando RL, Jannink J-L. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. *Genetics*. 2009;182(1):355–64.
4. Goddard ME, Hayes BJ. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat Rev Genet*. 2009;10(6):381–91.
5. Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME. Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genet Sel Evol*. 2009;41:51.
6. Song H, Zhang J, Jiang Y, Gao H, Tang S, Mi S, et al. Genomic prediction for growth and reproduction traits in pig using an admixed reference population. *J Anim Sci*. 2017;95(8):3415–24.
7. Song H, Ye S, Jiang Y, Zhang Z, Zhang Q, Ding X. Using imputation-based whole-genome sequencing data to improve the accuracy of genomic prediction for combined populations in pigs. *Genet Sel Evol*. 2019;51(1):58.
8. Lund MS, Roos AP, Vries AG, Druet T, Ducrocq V, Fritz S, et al. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet Sel Evol*. 2011;43:43.
9. Falconer DS, Mackay TF, Frankham R. Introduction to quantitative genetics (4th edn). *Trends Genet*. 1996;12(7):280.
10. Meyer K. Factor-analytic models for genotypex environment type problems and structured covariance matrices. *Genet Sel Evol*. 2009;41(1):21.
11. Liu A, Su G, Hoglund J, Zhang Z, Thomassen J, Christiansen I, et al. Genotype by environment interaction for female fertility traits under conventional and organic production systems in Danish Holsteins. *J Dairy Sci*. 2019;102(9):8134–47.
12. Kolmodin R, Strandberg E, Madsen P, Jensen J, Jorjani H. Genotype by environment interaction in Nordic dairy cattle studied using reaction norms. *Acta Agriculturae Scandinavica, Section A-Animal Sci*. 2002;52(1):11–24.
13. Su G, Madsen P, Lund MS, Sorensen D, Korsgaard IR, Jensen J. Bayesian analysis of the linear reaction norm model with unknown covariates. *J Anim Sci*. 2006;84(7):1651–7.
14. Zhou L, Ding XD, Zhang Q, Wang YC, Lund MS, Su GS. Consistency of linkage disequilibrium between Chinese and Nordic Holsteins and genomic prediction for Chinese Holsteins using a joint reference population. *Genet Sel Evol*. 2013;45:7.
15. Madsen P, Milkevyich V, Gao H, Christensen OF, Jensen J: DMU - A Package for Analyzing Multivariate Mixed Models in Quantitative Genetics and Genomics. In: Proceedings of the World Congress on Genetics Applied to Livestock Production. vol. Electronic Poster Session - Methods and Tools - Software; 2018: 525.

16. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet.* 2009;84(2):210–23.
17. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88(1):76–82.
18. Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet.* 1968;38(6):226–31.
19. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci.* 2008;91(11):4414–23.
20. Garrick D, Dekkers J, Fernando R. The evolution of methodologies for genomic prediction. *Livest Sci.* 2014;166:10–8.
21. Cheng H, Kizilkaya K, Zeng J, Garrick D, Fernando R. Genomic prediction from multiple-trait Bayesian regression methods using mixture priors. *Genetics.* 2018;209(1):89–103.
22. Cheng H, Fernando R, Garrick D: JWAS: Julia implementation of Whole-genome Analyses Software. In: *Proceedings of the World Congress on Genetics Applied to Livestock Production. vol. Methods and Tools - Software*; 2018: 859.
23. Robertson. The sampling variance of the genetic correlation coefficient. *Biometrics.* 1959;15:469–8.
24. Zhang Z, Kargo M, Liu AX, Thomassen JR, Pao YC, Su GS. Genotype-by-environment interaction of fertility traits in Danish Holstein cattle using a single-step genomic reaction norm model. *Heredity.* 2019;123(2):202–14.
25. Li X, Lund MS, Zhang Q, Costa CN, Ducrocq V, Su G. Short communication: improving accuracy of predicting breeding values in Brazilian Holstein population by adding data from Nordic and French Holstein populations. *J Dairy Sci.* 2016;99(6):4574–9.
26. Hay EH, Roberts A. Genotypex prenatal and post-weaning nutritional environment interaction in a composite beef cattle breed using reaction norms and a multi-trait model. *J Anim Sci.* 2018;96(2):444–53.
27. Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J Dairy Sci.* 2012;95(7):4114–29.
28. van den Berg I, Bowman PJ, MacLeod IM, Hayes BJ, Wang T, Bolormaa S, et al. Multi-breed genomic prediction using Bayes R with sequence data and dropping variants with a small effect. *Genet Sel Evol.* 2017;49(1):70.
29. De Roos A, Hayes BJ, Spelman R, Goddard ME. Linkage disequilibrium and persistence of phase in Holstein–Friesian. *Jersey Angus Cattle Genetics.* 2008;179(3):1503–12.
30. Hoze C, Fritz S, Phocas F, Boichard D, Ducrocq V, Croiseau P. Efficiency of multi-breed genomic selection for dairy cattle breeds with different sizes of reference population. *J Dairy Sci.* 2014;97(6):3918–29.
31. Li XJ, Lund MS, Janss L, Wang CL, Ding XD, Zhang Q, et al. The patterns of genomic variances and covariances across genome for milk production traits between Chinese and Nordic Holstein populations. *BMC Genet.* 2017;18:26.
32. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: genomic selection in dairy cattle: progress and challenges. *J Dairy Sci.* 2009;92(2):433–43.
33. Zhang CY, Kemp RA, Stothard P, Wang ZQ, Boddicker N, Krivushin K, et al. Genomic evaluation of feed efficiency component traits in Duroc pigs using 80K, 650K and whole-genome sequence variants. *Genet Sel Evol.* 2018;50:14.
34. Jia Y, Jannink JL. Multiple-Trait Genomic Selection Methods Increase Genetic Value Prediction Accuracy. *Genetics.* 2012;192(4):1513.
35. Misztal I, Legarra A, Aguilar I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J Dairy Sci.* 2009;92(9):4648–55.
36. Christensen OF, Lund MS. Genomic prediction when some animals are not genotyped. *Genet Sel Evol.* 2010;42:2.
37. Fernando RL, Dekkers JC, Garrick DJ. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genet Sel Evol.* 2014;46(1):50.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

