OXFORD

Briefings in Bioinformatics, 2023, **24(1)**, 1–10
https://doi.org/10.1093/bib/bbac587
Problem Solving Protocol

Development and application of an efficient genomic mating method to maximize the production performances of three-way crossbred pigs

Zhenshuang Tang†, Lilin Yin†, Dong Yin, Haohao Zhang, Yuhua Fu, Guangliang Zhou, Yunxiang Zhao, Zhiquan Wang, Xiaolei Liu, Xinyun Li and Shuhong Zhao

Corresponding authors. Xiaolei Liu, Animal Genetics and Breeding, College of Animal Science and Technology, Huazhong Agricultural University. Box: 430070, Wuhan, PR China, Tel.: +86 18571857369; E-mail: xiaoleiliu@mail.hzau.edu.cn; Xinyun Li, Animal Genetics and Breeding, College of Animal Science and Technology, Huazhong Agricultural University. Box: 430070, Wuhan, PR China, Tel.: +86 15827306097; E-mail: xyli@mail.hzau.edu.cn; Shuhong Zhao, Animal Genetics and Breeding, College of Animal Science and Technology, Huazhong Agricultural University. Box: 430070, Wuhan, PR China, Tel.: +8618986120366; E-mail: shzhao@mail.hzau.edu.cn.

†Zhenshuang Tang and Lilin Yin contributed equally to this work.

Abstract

Creating synthetic lines is the standard mating mode for commercial pig production. Traditional mating performance was evaluated through a strictly designed cross-combination test at the 'breed level' to maximize the benefits of production. The Duroc–Landrace–Yorkshire (DLY) three-way crossbred production system became the most widely used breeding scheme for pigs. Here, we proposed an 'individual level' genomic mating procedure that can be applied to commercial pig production with efficient algorithms for estimating marker effects and for allocating the appropriate boar-sow pairs, which can be freely accessed to public in our developed HIBLUP software at https://www.hiblup.com/tutorials#genomic-mating. A total of 875 Duroc boars, 350 Landrace–Yorkshire sows and 3573 DLY pigs were used to carry out the genomic mating to assess the production benefits theoretically. The results showed that genomic mating significantly improved the performances of progeny across different traits compared with random mating, such as the feed conversion rate, days from 30 to 120 kg and eye muscle area could be improved by −0.12, −4.64 d and 2.65 cm^2, respectively, which were consistent with the real experimental validations. Overall, our findings indicated that genomic mating is an effective strategy to improve the performances of progeny by maximizing their total genetic merit with consideration of both additive and dominant effects. Also, a herd of boars from a richer genetic source will increase the effectiveness of genomic mating further.

Keywords: genomic mating, three-way crossbred pigs, feed efficiency, growth, carcass, dominant effect

Introduction

As the world's population grows, and extreme weather and epidemic outbreaks increase in frequency, food security is one of the biggest challenges that we face [1]. In addition, food competition between humans and animals is intensifying. Improving energy conversion efficiency from feed to animal products could mitigate this pressure [2]. Hybridization is the most common mode of livestock production, which utilizes additive and dominant effects from a combination of breeds to maximum the production performance of hybrid progeny.

In pigs, crossbreeding between the Duroc, Landrace and Yorkshire is the main approach for commercial pig production. Crossbred sows between the Yorkshire and Landrace breeds mate with terminal Duroc boars to capitalize on the genetic heterosis from

Zhenshuang Tang is a Ph.D. student of Animal Genetics and Breeding at the College of Animal Science and Technology, Huazhong Agricultural University, focusing on developing statistical methods for Genomic Mating and Genomic Selection.
Lilin Yin is a Postdoc at the Frontiers Science Center for Animal Breeding and Sustainable Production, College of Animal Science and Technology, Huazhong Agricultural University, focusing on developing algorithms and software tools for Genomic Selection and Genome Wide Association Study.
Dong Yin is a Ph.D. student of Animal Genetics and Breeding at the College of Animal Science and Technology, Huazhong Agricultural University, focusing on developing software tools for simulating complex breeding programs.
Haohao Zhang is a Ph.D. student of Computer Science and Technology at the School of Computer Science and Artificial Intelligence, Wuhan University of Technology, focusing on developing software tools for Genomic Selection and Genome Wide Association Study.
Yuhua Fu is a Postdoc at the Frontiers Science Center for Animal Breeding and Sustainable Production, College of Animal Science and Technology, Huazhong Agricultural University, focusing on developing multi-omics integration algorithms and databases.
Guangliang Zhou is a Ph.D. student of Animal Genetics and Breeding at the College of Animal Science and Technology, Huazhong Agricultural University, focusing on developing algorithms for Genome Wide Association Study.
Yunxiang Zhao is a Lecturer at the School of Life Sciences and Engineering, Foshan University, focusing on pig breeding.
Zhiquan Wang is an expert advisor at the Wuhan Yingzi Gene Technology Co. LTD. He is a statistical geneticist, working on quantitative genetics.
Xiaolei Liu is a professor at the Frontiers Science Center for Animal Breeding and Sustainable Production, College of Animal Science and Technology, Huazhong Agricultural University & Hubei Hongshan Laboratory, focusing on developing statistical algorithms and software tools for Genome Wide Association Study and Genomic Selection.
Xinyun Li is a professor at the Frontiers Science Center for Animal Breeding and Sustainable Production, College of Animal Science and Technology, Huazhong Agricultural University, working on pig genomic breeding.
Shuhong Zhao is a professor at the Frontiers Science Center for Animal Breeding and Sustainable Production, College of Animal Science and Technology, Huazhong Agricultural University & Hubei Hongshan Laboratory, focusing on pig genomic breeding.
Received: October 22, 2022. **Revised:** November 24, 2022. **Accepted:** November 30, 2022

dominant effects that is exhibited in their progeny, which is usually expressed with rapid growth and better feed conversion efficiency [3, 4].

Mating is the process of deciding which male and female should form a mating pair to optimize the performance of objective traits in their offspring generation, and it has a long history in animal breeding [5, 6]. Traditional mating is mainly to control inbreeding by pedigree information to avoid mating between genetically related animals. Destefano and Hoeschele [7] evaluated the effect of specific combining ability in mating and illustrated that dominant effects from pedigrees yielded progeny merits that were slightly but significantly higher than random mating using a simulation study. Hayes and Miller [8] developed a 'super model' to exploit across- and within-breed dominant variation simultaneously with the aim of maximizing the genetic merit of progeny, which showed that mate selection on total progeny merit improved progeny performance ∼12.5% over truncation selection when dominant variation was large relative to total genetic variance. Weigel and Lin [9] evaluated the outcome of sire selection and mate selection with three different objectives and found that mate selection reduced inbreeding in the next generation and led to an increase in farm profitability in Holstein and Jersey cattle.

By the year 2000, the information mainly from phenotypic and pedigree records was used in mate allocation, but mating did not function optimally unless extensive, historical pedigree data for capturing dominant variation were available. With the development of sequencing technology over the past decade, high-density genetic markers across the genome could be obtained, and genomic information was utilized in mate allocation. Different from traditional mating, genomic mating utilizes genomic information to design mating pairs to control inbreeding and to maximize progeny performance, this method has been conducted on dairy cattle and pigs [10, 11].

Note that any improvement due to heterosis cannot be inherited over generations because heterosis is a specific combining ability [12]. Toro and Varona [13] quantified the efficiency of genomic mating using simulated data, and they showed that mate allocation that included both additive and dominant effects provided an additional genetic merit that ranged from 6 to 22%. Aliloo *et al*. [11] proposed a general framework for genomic mating that involved additive, dominant effects and average heterozygosity using real Holstein cattle data, and their results indicated that an average of 8.42 Australian dollars more profit per mating could be achieved by genomic mating. Gonzalez-Dieguez *et al*. [10] conducted genomic mating based on total genetic values on purebred Landrace pigs and showed that genomic mating improved progeny performance by −0.79 d and −0.04 mm for age at 100 kg and backfat thickness, respectively. Until now, genomic mating has not been carried out for parents of a commercial pig population for feed efficiency, growth and carcass traits.

DLY three-way crossbred production is one of the most efficient strategies to produce commercial pigs. The genomic mating is a potential strategy to improve the performance of commercial pigs further by switching the mating between 'breeds' and 'individuals'. That is, for any Landrace–Yorkshire sow, we can use the additive and dominant effects of whole genome markers to select the most suitable Duroc boar to replace a boar that was selected randomly. The aims of this study were: (i) to construct a genomic mating procedure in a typical DLY three-way crossbred production system; (ii) to develop an efficient algorithm to allocate a suitable pair of a Landrace–Yorkshire sow and a Duroc boar to maximize the production performance of DLY three-way crossbred pigs and (iii) to evaluate the additional production benefits brought by genomic mating compared with random mating.

Materials and methods

Simulated data

A total of 300 Duroc boars and 30 Landrace–Yorkshire crossbred sows with real genotypic data, which included 47 100 single nucleotide polymorphisms (SNPs), were used to simulate the genotype and phenotype of DLY three-way crossbred pigs using Simer (https://github.com/xiaolei-lab/SIMER). An ideal fixed mating mode was utilized to reproduce the progeny, where each boar mated once with each sow, which resulted in 9000 mating pairs. To facilitate the comparison of simulation results and practical experimental results, assuming the litter size was 14, there were finally 126 000 three-way crossbred pigs in total. One trait was simulated for the three-way crossbred pigs, the additive effect controlled by randomly selected 1000 quantitative trait nucleotides (QTNs) had a zero mean and a variance of 30, and the additive heritability was 0.3; the dominant effect controlled by randomly selected 100 QTNs had a zero mean and a variance of 15, and the dominant heritability was 0.15; the residual variance was set to 55. The average phenotypic values and average total genetic values (TGV) of each mating pair were computed based on the additive genetic values, dominant genetic values and residuals of pigs within each mating pair.

Real data

Phenotype collection

The phenotypes of DLY three-way crossbred pigs, which included feed efficiency (feed conversion rate (FCR), average daily feed intake (ADFI)), growth (days from 30 to 120 kg (AGE), average daily gain (ADG)), and carcass (backfat thickness (BFT), eye muscle area (EMA)) traits, were collected from a typical commercial pig population (Supplementary Note). The feed intake and body weight were obtained by the Nedap Pig Performance Testing (PPT) Station, and the pigs were measured beginning at 30 or 60 kg, and all pigs were off test at 115–120 kg. The errors in feed intake records were corrected using eight criteria proposed by Casey *et al*. [14]. The body weight data were obtained by fitting a robust regression following the procedures of Jiao *et al*. [15].

Genotyping

Genomic DNA was extracted from frozen ear or tail tissue samples. The PorcineSNP50 Bead Chip was used for genotyping. The number of genotyping pigs is shown on Table 1, and all SNP markers were mapped to *Sus scrofa* genome build 11.1.

Missing genotype values were fully imputed using Beagle 5.2 software [16]. For three-way crossbred pigs, data quality control was performed using PLINK [17] and G2P [18] software before and after imputation. The SNPs genotype call rate < 0.9 and minor allele frequency < 0.01 were filtered out for the imputed data. After processing, 47 100 SNPs remained for subsequent data analysis (Supplementary Figure S1, see Supplementary Data available online; [19]).

Estimation of variance components

For each trait, four types of linear mixed models with varied fixed effects and random effects were fitted to estimate the variance components.

Table 1. Numbers of Duroc, two-way crossbred and three-way crossbred pigs used in genomic mating

Type	Sex	Numbers of genotyped pigs
Duroc	Male	3200
Two-way crossbred	Female	350
Three-way crossbred	Male/Female	4857

Model 1 only considers additive effects:

$$y = Xb + Za + e$$

Model 2 includes additive and dominant effects:

$$y = Xb + Za + Wd + e$$

Model 3 includes additive effects and average heterozygosity:

$$y = Xb + Hf + Za + e$$

Model 4 takes additive, dominant effects and average heterozygosity into account:

$$y = Xb + Hf + Za + Wd + e$$

where y is the vector of phenotypic observations for each trait; X is the incidence matrix for fixed effects, b is the corresponding effects vector; H is the average heterozygosity of each animal as described by Aliloo *et al.* [11]:

$$H_i = \frac{\sum_{k=1}^{m} h_{ki}}{\sum_{k=1}^{m} 2p_k(1-p_k)}$$

where H_i is the average heterozygosity of individual i across all the markers, h_{ki} is the corresponding element of dominant marker covariates matrix for individual i at the k**th** marker, m is the number of markers and p_k is the frequency of allele A at the k**th** marker; f is the regression coefficient on average heterozygosity of animals; Z and W are the incidence matrix that relates observations to additive and dominant random effects; a is the vector of breeding values of animals, it is commonly assumed to be distributed as $a \sim N(0, G_A\sigma_A^2)$, where G_A is the additive genomic relationship matrix (GRM), and σ_A^2 is the additive genetic variance; d is the vector of dominant deviations of animals, which is distributed as $d \sim N(0, G_D\sigma_D^2)$, where G_D is the dominant GRM, and σ_D^2 is the dominant variance; e is the vector of random residual terms following the normal distribution $N(0, I\sigma_e^2)$, where I is the identity matrix, and σ_e^2 is the residual variance.

The G_A and G_D were constructed based on genome-wide markers as defined in Vitezica *et al.* [20]:

$$G_A = \frac{M_A M_A^T}{tr(M_A M_A^T)/n}$$

$$G_D = \frac{M_D M_D^T}{tr(M_D M_D^T)/n}$$

where n is the number of individuals; M_A is the additive marker covariates matrix as follows:

$$M_A = \begin{cases} -(-p_{Aa} - 2p_{aa}) \\ -(1 - p_{Aa} - 2p_{aa}) \\ -(2 - p_{Aa} - 2p_{aa}) \end{cases} \text{ for genotypes } \begin{cases} AA \\ Aa \\ aa \end{cases}$$

M_D is the dominant marker covariates matrix as:

$$M_D = \begin{cases} -\frac{2p_{Aa}p_{aa}}{p_{AA}+p_{aa}-(p_{AA}-p_{aa})^2} \\ \frac{4p_{AA}p_{aa}}{p_{AA}+p_{aa}-(p_{AA}-p_{aa})^2} \\ -\frac{2p_{Aa}p_{AA}}{p_{AA}+p_{aa}-(p_{AA}-p_{aa})^2} \end{cases} \text{ for genotypes } \begin{cases} AA \\ Aa \\ aa \end{cases}$$

where p_{AA}, p_{Aa} and p_{aa} are the genotypic frequencies for the genotypes AA, Aa and aa, respectively.

The variance components and heritability were estimated using the average information restricted maximum likelihood (AI-REML) algorithm [21] in HIBLUP software (https://www.hiblup.com/). For the real experimental data, the Likelihood Ratio Test (LRT) algorithm and Akaike Information Criterion (AIC) statistic was used to select the random effects and fixed effects in model, and the factor 'test station' was finally tested significantly to be fitted in model as an environmental random effect.

Estimation of Marker effects

The most commonly used algorithm to estimate additive and dominant effects is to derive it from the prior calculated genetic random effects as follows [11, 22]:

$$\beta_A = \frac{M_A^T G_A^{-1} a}{tr\left(M_A M_A^T\right)/n}$$

$$\beta_D = \frac{M_D^T G_D^{-1} d}{tr\left(M_D M_D^T\right)/n}$$

where β_A and β_D are the estimated additive and dominant effects, respectively. The a and d are the additive and dominant genetic random effects, which can be obtained by solving the following mixed model equation (MME),

$$\begin{bmatrix} X^TX & X^TZ & X^TW \\ Z^TX & Z^TZ + \frac{\sigma_e^2}{\sigma_A^2}G_A^{-1} & Z^TW \\ W^TX & W^TZ & W^TW + \frac{\sigma_e^2}{\sigma_D^2}G_D^{-1} \end{bmatrix} \begin{bmatrix} b \\ a \\ d \end{bmatrix} = \begin{bmatrix} X^Ty \\ Z^Ty \\ W^Ty \end{bmatrix}$$

As shown above, the whole processes of calculating marker effects involve multiple times of inverse computations for several big matrices including the additive and dominant GRM, and the coefficient matrix of MME. Moreover, the dimension of MME is at least two times bigger than the total number of individuals with and without phenotypic records for the model in which there are only two random effects, and the ratio appears to have linear growth with more random effects fitted in the model. Therefore, it is expensive on either time spending or memory consumption for handling big genomic breeding data. Here, we proposed an efficient algorithm for calculating marker effects, it can be formulated as follows:

$$V = ZG_AZ^T\sigma_A^2 + WG_DW^T\sigma_D^2 + I\sigma_e^2$$

$$P = V^{-1} - V^{-1}X\left(X^{T}V^{-1}X\right)^{-1}X^{T}V^{-1}$$

$$\beta_{A} = \frac{M_{A}^{T}Z^{T}Py\sigma_{A}^{2}}{tr\left(M_{A}M_{A}^{T}\right)/n}$$

$$\beta_{D} = \frac{M_{D}^{T}W^{T}Py\sigma_{D}^{2}}{tr\left(M_{D}M_{D}^{T}\right)/n}$$

all elements in the above formulas are the same as defined previously. Unlike the MME-based algorithm, this strategy does not require to solve MME to get the breeding values in prior. No matter how many random effects are included in the model, the dimension of V matrix remains unchanged and it is equal to the number of individuals with effective phenotypic records, which is generally smaller than the dimension of GRM and MME. In addition, the whole processes only involve computing the inverse of V matrix instead of all GRM, making it competent in handling a large model with numerous random effects. However, the core part Py can be derived efficiently from solving the following system of linear equations:

$$V[\phi_1 \quad \phi_2] = [X\ y]$$

$$Py = \phi_2 - \phi_1\left(X^{T}\phi_1\right)^{-1}X^{T}\phi_2$$

the solution ϕ_1 and ϕ_2 can be obtained efficiently using either factorization algorithms (e.g. cholesky and lu decomposition) or iterative algorithms (e.g. preconditioned conjugate gradient), thus it is not required to compute the inverse of V matrix directly. Since the number of fixed effects is far smaller than the number of individuals in general, computing the inverse of $X^{T}\phi_1$ could be accomplished on-the-fly. Owing to lower dimension of matrix involved in operations and no requirements of computing inverse directly for any big matrix, the proposed strategy above should be more computationally beneficial regarding the time spending and memory consumption in the analysis of big genomic breeding data while generating exactly the same results with MME-based strategy.

Predicting expected progeny value

The genotypic values of progeny at locus k were computed by the following equations:

$$\begin{cases} \hat{g_{AA}} = (2 - 2p_k) \times \beta_A + \left(-2q_k^2\right) \times \beta_D \\ \hat{g_{Aa}} = (1 - 2p_k) \times \beta_A + (2p_k q_k) \times \beta_D \\ \hat{g_{aa}} = (0 - 2p_k) \times \beta_A + \left(-2p_k^2\right) \times \beta_D \end{cases}$$

where $\hat{g_{AA}}$, $\hat{g_{Aa}}$ and $\hat{g_{aa}}$ are the genotypic values for AA, Aa and aa genotype, respectively; β_A is the estimated additive effects; β_D is the estimated dominant effects, when mating does not take into account dominant effects, β_D will be equal to zero **and** p_k is the frequency of allele A at the k^{th} marker, $q_k = 1 - p_k$.

The expected total genetic value of progeny, which was also termed expected progeny value (EPV) by Aliloo *et al.* [11], was calculated as the weighted average of expected genetic value as follows:

$$\hat{EPV} = \sum_{k=1}^{m}\left[p(AA)\hat{g_{AA}} + p(Aa)\hat{g_{Aa}} + p(aa)\hat{g_{aa}}\right]$$

where $\hat{EPV}$ is the expected progeny value from a mating between i^{th} Duroc boar and j^{th} Landrace–Yorkshire crossbred sow, which was computed by summing the three genotypic values time their corresponding genotype probabilities in the probability of expected genotype distribution table (Supplementary Table S1); $p(AA)$ is the genotype probability and $\hat{g_{AA}}$ is the genotypic value at the k**th** locus with three possible genotypes of AA, Aa and aa; **and** m is the number of markers.

Genomic mating procedure

For the simulated data, the genotype data of 300 boars and 30 crossbred sows and the estimated marker effects derived from their offspring were used to carry out genomic mating based on total genetic values and breeding values, and the EPV of each mating pair was obtained. The outcomes of genomic mating of simulated data were evaluated using the correlation between EPV and phenotypic values or between EPV and TGV.

For the real data, 875 Duroc boars and 350 Landrace–Yorkshire crossbred sows were used to perform genomic mating according to the procedure shown in Figure 1. Positive selection was carried out for ADG and EMA, and negative selection was implemented for FCR, AGE, ADFI and BFT. Based on the genomic mating module in the HIBLUP software, EPV of each mating pair was calculated, and the optimal mating solution was obtained by linear programming. There were two constraint conditions when solving the optimal mating pair: each boar only could be mated with 20 sows, and each sow could not be allocated to more than one boar. The linear function can be written as:

$$f_{optimal}\left(\hat{EPV}_{ij}\right) = \sum_{i=1}^{b}\sum_{j=1}^{s}\hat{EPV}_{ij}x_{ij}$$

where the constraint condition of $x_{i1} + x_{i2} + \cdots + x_{is} = 20\ (i = 1, 2, \ldots, b)$ is for boar i, and the $x_{1j} + x_{2j} + \cdots + x_{bj} = 1\ (j = 1, 2, \ldots, s)$ is for sow j; b and s are the total number of boars and sows; x_{ij} equals 0 or 1, where 0 means that the mating between boar i and sow j was not selected, and 1 indicates the corresponding mating was selected. The mean EPV of 100 times sampled randomly with constraint conditions was viewed as random mating. The outcome of genomic mating was the difference in EPV between genomic mating (GM) and random mating (RM).

Experimental validation of genomic mating

A designed experiment was carried out to validate and to evaluate the efficiency of genomic mating with the GM group and the RM group. In the GM group, five candidate boars were listed for each sow based on their EPV rankings. The boars in the RM group were selected randomly at the same boar station. The number of boars and sows used in the experiment were summarized on Supplementary Table S2. However, due to the different estrous time of sows and the semen availabilities of boars, the experimental results were different from those under ideal conditions. The DLY pigs, which are the progeny of experimental sows, were reared under the same environmental and feed rhythm. Fourteen pigs were fed in a feeding station, in which half of them came from the GM group, and the other half from the RM group. Pigs in the same feeding station had similar birth dates and body weights. The summary information of the experimental DLY pigs was shown in Supplementary Table S3, and their performances were measured using the Nedap PPT system. The quality control method described in *Phenotype collection* was used to process the phenotypic records. The phenotype differences between the GM and the RM groups were analyzed by Student's t-test.

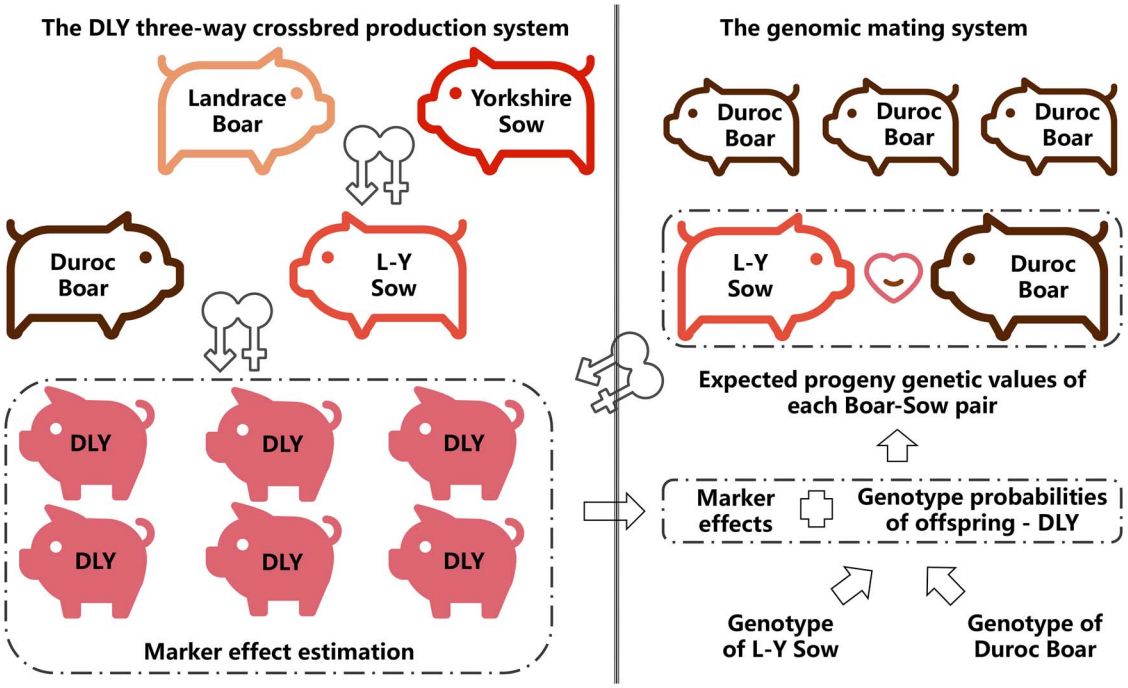

Figure 1. Genomic mating procedure. The genomic mating procedure of Duroc–Landrace–Yorkshire (DLY) three-way crossbred pigs is shown. The genotype and phenotype of the DLY population were used to estimate the marker effects. The genotypes of Landrace–Yorkshire (L–Y) sow and Duroc boar were used to calculate the genotype probabilities of offspring for all possible mating pairs. Expected progeny genetic values were calculated by multiplying the genotype probabilities by the genotypic values, which is used as the basis for mate allocation.

Table 2. Summary statistics of traits in the three-way crossbred pig population

Trait[a]	Number	Max	Min	Median	Mean	SD[b]	CV[c]
FCR	3573	4.60	1.84	2.74	2.76	0.30	0.11
ADFI (g)	3573	3801.88	1772.30	2691.42	2695.91	308.20	0.11
AGE (days)	3573	183.55	62.56	91.17	93.34	14.39	0.15
ADG (g)	3573	1438.54	490.32	987.20	985.46	141.25	0.14
BFT (mm)	2318	23.60	4.30	12.40	12.57	2.71	0.22
EMA (cm^2)	2319	62.90	13.61	37.89	38.09	5.74	0.15

[a]FCR, feed conversion rate; ADFI, average daily feed intake; AGE, days from 30 to 120 kg; ADG, average daily gain; BFT, backfat thickness; EMA, eye muscle area. [b]SD, standard deviation. [c]CV, variable coefficient.

Factors that affected genomic mating

In this study, we evaluated the effect of statistical models and boar population size on genomic mating performances. Four models were used to estimate the additive and dominant effects. The boars were selected randomly from a boar station and set up in six different groups from 500 to 3000, which increased by 500 each time. All other factors were held constant when evaluating these two factors.

Results

Summary information on phenotypic data

The growth curve of one pig was selected randomly, and its body weight against age that removed the abnormal values is shown in Supplementary Figure S2. All traits were distributed normally (Table 2), and the quantile–quantile (QQ) plots are shown in Supplementary Figure S3.

Estimation of variance components and heritabilities

The genetic variances were calculated by using both simulated and real data. The results of simulated data (Supplementary Table S4) indicated that the variance components were estimated accurately using Vitezica *et al.*'s [20] method. Also, the trend of estimated genetic parameters of the four models between simulated data and real data were similar.

For the real data, the estimated additive heritability ranged from 0.2 to 0.4, the estimated dominant heritability was between 0.1 and 0.2 and the estimated ratio of σ_D^2/σ_A^2 ranged from 0.4 to 0.8 (Table 3). The estimated variance components varied among the four models. When dominant effects were considered in models 2 and 4, both additive and residual variances decreased compared with that of models 1 and 3, which suggested that the residual variances cannot capture the main part of dominant variances, some of which came from additive effects. The effect of average heterozygosity on the estimation of variance components was not significant because estimates of models 1 and 3 were similar, and models 2 and 4 were also similar to each other.

The effect of genomic mating

There were similarities between the simulated data and the real data (Table 4 and Supplementary Table S5). The EPVs from the model based on total genetic values were highest. There were similar EPVs as in models 1 and 3, which suggested that the average heterozygosity did not affect mate allocation

Table 3. Variance components, heritabilities and heterozygosity effects obtained from four marker effects estimation models

Trait[a]	Model[b]	Var_test.station (±SE)[c]	Var_GA (±SE)[c]	Var_GD (±SE)[c]	Var_e (±SE)[c]	Het (±SE)[c]	h^2_GA (±SE)[c]	h^2_GD (±SE)[b]	Var_GD/Var_GA
FCR	1	0.0068 (0.0012)	0.0303 (0.0026)		0.0483 (0.0017)		0.3548 (0.0243)		
	2	0.0067 (0.0012)	0.0239 (0.0025)	0.0110 (0.0025)	0.0418 (0.0022)		0.2865 (0.0265)	0.1322 (0.0294)	0.4603
	3	0.0068 (0.0012)	0.0293 (0.0026)		0.0486 (0.0017)	−1.0694 (0.3785)	0.3463 (0.0244)		
	4	0.0067 (0.0012)	0.0240 (0.0025)	0.0108 (0.0025)	0.0420 (0.0022)	−0.4753 (0.5029)	0.2871 (0.0265)	0.1292 (0.0297)	0.4500
ADFI	1	5173.8864 (1004.4757)	32 091.6995 (2756.3684)		51 583.2416 (1798.3891)		0.3612 (0.0246)		
	2	5116.0309 (992.5801)	19 740.3660 (2457.7211)	15 342.3516 (2608.7979)	43 885.0072 (2233.2978)		0.2348 (0.0263)	0.1825 (0.0304)	0.7772
	3	5154.2718 (1001.6154)	29 504.4490 (2668.4274)		52 354.4488 (1808.4153)	−2087.8163 (387.7728)	0.3391 (0.0249)		
	4	5111.6140 (992.0321)	20 042.0376 (2470.8290)	14 560.1680 (2632.8594)	44 290.1104 (2255.7724)	−932.8431 (535.1313)	0.2386 (0.0264)	0.1733 (0.0308)	0.7265
AGE	1	17.0783 (3.0701)	55.0444 (5.7604)		133.9107 (4.4467)		0.2672 (0.0242)		
	2	16.9507 (3.0538)	43.1357 (5.7415)	23.9109 (6.1971)	119.4816 (5.6265)		0.2120 (0.0258)	0.1175 (0.0302)	0.5543
	3	17.0771 (3.0704)	54.4855 (5.7390)		134.0003 (4.4468)	44.0267 (18.6079)	0.2651 (0.0242)		
	4	16.9484 (3.0536)	43.5582 (5.7667)	23.2423 (6.2414)	119.7723 (5.6524)	27.3566 (24.2947)	0.2140 (0.0259)	0.1142 (0.0305)	0.5336
ADG	1	1753.2180 (308.0495)	5804.6677 (574.6649)		12 377.3718 (419.2357)		0.2912 (0.0245)		
	2	1747.4619 (307.4327)	4580.8502 (572.5291)	2342.2535 (594.9440)	10 990.7175 (531.8382)		0.2330 (0.0263)	0.1191 (0.0301)	0.5113
	3	1752.8818 (308.0711)	5747.7704 (572.7513)		12 392.8985 (419.4747)	−379.3138 (182.8085)	0.2889 (0.0245)		
	4	1747.0091 (307.3961)	4608.2345 (574.1890)	2301.6139 (598.9491)	11 009.3299 (533.9686)	−213.9463 (239.4711)	0.2343 (0.0264)	0.1170 (0.0303)	0.4995
BFT	1	0.2850 (0.0812)	2.2785 (0.2696)		4.5445 (0.1998)		0.3206 (0.0317)		
	2	0.3061 (0.0840)	1.3574 (0.2453)	0.9397 (0.2488)	4.1823 (0.2422)		0.2000 (0.0334)	0.1385 (0.0362)	0.6923
	3	0.2969 (0.0830)	2.0862 (0.2632)		4.6225 (0.2016)	−14.0128 (4.2042)	0.2978 (0.0320)		
	4	0.3078 (0.0843)	1.3657 (0.2461)	0.9106 (0.2546)	4.2014 (0.2452)	−3.6924 (5.2155)	0.2013 (0.0335)	0.1342 (0.0371)	0.6668
EMA	1	3.0336 (0.6139)	8.2177 (1.0681)		20.0029 (0.8535)		0.2629 (0.0302)		
	2	3.0141 (0.6107)	7.0701 (1.1096)	3.2242 (1.3313)	17.8911 (1.1662)		0.2266 (0.0327)	0.1033 (0.0424)	0.4560
	3	3.0194 (0.6119)	8.1834 (1.0669)		20.0156 (0.8538)	−12.3771 (8.6496)	0.2621 (0.0302)		
	4	3.0113 (0.6105)	7.1014 (1.1124)	3.1340 (1.3481)	17.9527 (1.1764)	−7.5507 (10.5800)	0.2276 (0.0328)	0.1005 (0.0429)	0.4413

[a]FCR, feed conversion rate; ADFI, average daily feed intake; AGE, days from 30 to 120 kg; ADG, average daily gain; BFT, backfat thickness; EMA, eye muscle area. [b]Represents four models for marker effects estimation. 1: only considers additive effects; 2: includes both additive and dominant effects; 3: includes additive effects and average heterozygosity and 4: takes additive, dominant effects and average heterozygosity into account. [c]Var_test.station: the variance of test station; SE: the standard error; Var_GA: the additive genetic variance; Var_GD: the dominant genetic variance; Var_e: the residual variance; Het: the average heterozygosity effect; h2_GA: the additive heritability; h2_GD: the dominant heritability.

Table 4. The average expected progeny values for six traits from a typical DLY pigs

Mating model	Effect model[b]	FCR (±SE)[a]	ADFI, g (±SE)[a]	AGE, days (±SE)[a]	ADG, g (±SE)[a]	BFT, mm (±SE)[a]	EMA, cm^2 (±SE)[a]
Breeding values	1	−0.1164 (0.0037)	−126.6165 (4.2195)	−4.2588 (0.1207)	43.8579 (1.2337)	−1.0760 (0.0324)	2.5619 (0.0662)
	2	−0.1013 (0.0031)	−93.9488 (3.1280)	−3.5922 (0.1000)	37.3971 (1.0514)	−0.8077 (0.0237)	2.2482 (0.0579)
	3	−0.1166 (0.0036)	−124.3126 (3.9386)	−4.2516 (0.1239)	43.9737 (1.2661)	−1.0288 (0.0315)	2.5644 (0.0653)
	4	−0.1018 (0.0031)	−95.3136 (3.1642)	−3.6157 (0.0997)	37.6254 (1.0347)	−0.8100 (0.0241)	2.2645 (0.0572)
Total genetic values	1	−0.1164 (0.0037)	−126.6165 (4.2195)	−4.2588 (0.1207)	43.8579 (1.2337)	−1.0760 (0.0324)	2.5619 (0.0662)
	2	−0.1222 (0.0035)	−132.1823 (3.4042)	−4.6783 (0.1080)	48.3496 (1.1145)	−1.0199 (0.0237)	2.6502 (0.0614)
	3	−0.1166 (0.0036)	−124.3126 (3.9386)	−4.2516 (0.1239)	43.9737 (1.2661)	−1.0288 (0.0315)	2.5644 (0.0653)
	4	−0.1220 (0.0034)	−130.4313 (3.4100)	−4.6409 (0.1085)	48.2637 (1.1085)	−1.0162 (0.0240)	2.6515 (0.0601)

[a]FCR, feed conversion rate; SE represents the standard error; ADFI, average daily feed intake; AGE, days from 30 to 120 kg; ADG, average daily gain; BFT, backfat thickness; EMA, eye muscle area. [b]Represents four models for marker effects estimation. 1: only considers additive effects; 2: includes both additive and dominant effects; 3: includes additive effects and average heterozygosity and 4: takes additive, dominant effects and average heterozygosity into account.

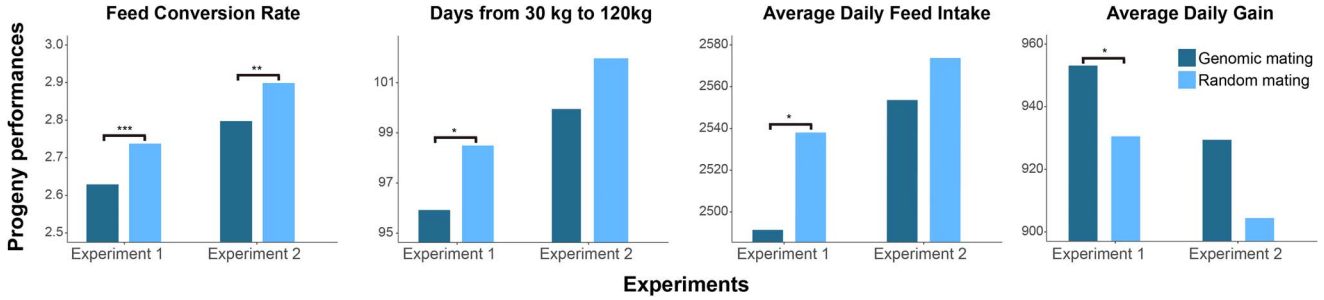

Figure 2. Results of validation experiment of genomic mating. Two separate experiments were carried out to validate the effect of genomic mating. The progeny performances of four traits in the genomic mating group and the random mating group were compared in each experiment. The *y*-axis represented the progeny performances and *x*-axis represented the two experiments. The asterisks (* $P < 0.05$,** $P < 0.01$ and ***$P < 0.001$) represented that the progeny performance was significantly different between the genomic mating group and the random mating group.

significantly. The results of genomic mating based on breeding values showed that marker effects from models 1 and 3 were better than those from models 2 and 4. In contrast, the outcomes of mating based on total genetic values showed that the marker effects from models 2 and 4 were better than those from models 1 and 3. These findings indicated that a portion of the dominant effects was separated from additive effects as well, and including dominant effects could help to improve genomic mating performances.

Compared with random mating, genomic mating improved the performances of progeny for FCR, ADFI, AGE, ADG, BFT and EMA by −0.12, −130.43 g, −4.64 d, 48.26 g, −1.02 mm and 2.65 cm^2, respectively. The progeny performances from genomic mating allocation based on the total genetic values calculated by model 4 improved 19.84, 36.84, 28.35, 28.27, 25.46 and 17.09% for FCR, ADFI, AGE, ADG, BFT and EMA, respectively, compared with using additive genetic values only.

Validation of genomic mating

The results of real experiments were consistent with the theoretical analysis. Compared with random mating, the implementation of genomic mating improved FCR by 0.1 point and shortened AGE by >2 d (Figure 2). The experimental results given separately by the sex were in rough agreement with using all progenies, but the performances of castrated boars and sows were different on some of the growth traits, such as ADFI and AGE (Supplementary Tables S6–S8).

Boar population size and models for estimating marker effects

With the increase in the boars' population size, the genetic pool became more abundant, the likelihood of a sow mating with the best boar increased and the effect of genomic mating was better (Figure 3 and Supplementary Table S9). The marker effects estimated by the multivariate model were used to compute the EPVs for FCR and AGE, and it was generally consistent with the univariate model (Supplementary Table S10). The additive and dominant genetic correlations of traits were also estimated using the multivariate model (Supplementary Tables S11 and S12).

Discussion

Genomic selection and genomic mating can be carried out simultaneously in a three-way hybrid breeding and production system to maximize both the breeding efficiency and production efficiency. This is first study to systematically explore the genomic mating and to implement genetic evaluation using real genotype and phenotype data from thousands of individuals for production traits in a typical DLY three-way crossbred commercial pig population. A complete genomic mating procedure with an efficient algorithm for estimating marker effects was developed to select the appropriate boar-sow pair to produce progeny. The evaluation results of theoretical analysis showed that the performances of progeny from mating pairs in genomic mating improved FCR by −0.12 point and AGE by −4.64 d compared with random mating. On average, to incorporate both additive and dominant effects improved the progeny performances by 25.98% compared with incorporating additive effects only, which means that genetic heterosis from dominant effects can be used to maximize the production performances of progeny in mate allocation.

Efficient algorithm for estimating marker effects

The algorithm for estimating marker effects was first used in mate allocation. Unlike existing methods [10, 11, 23], our strategy

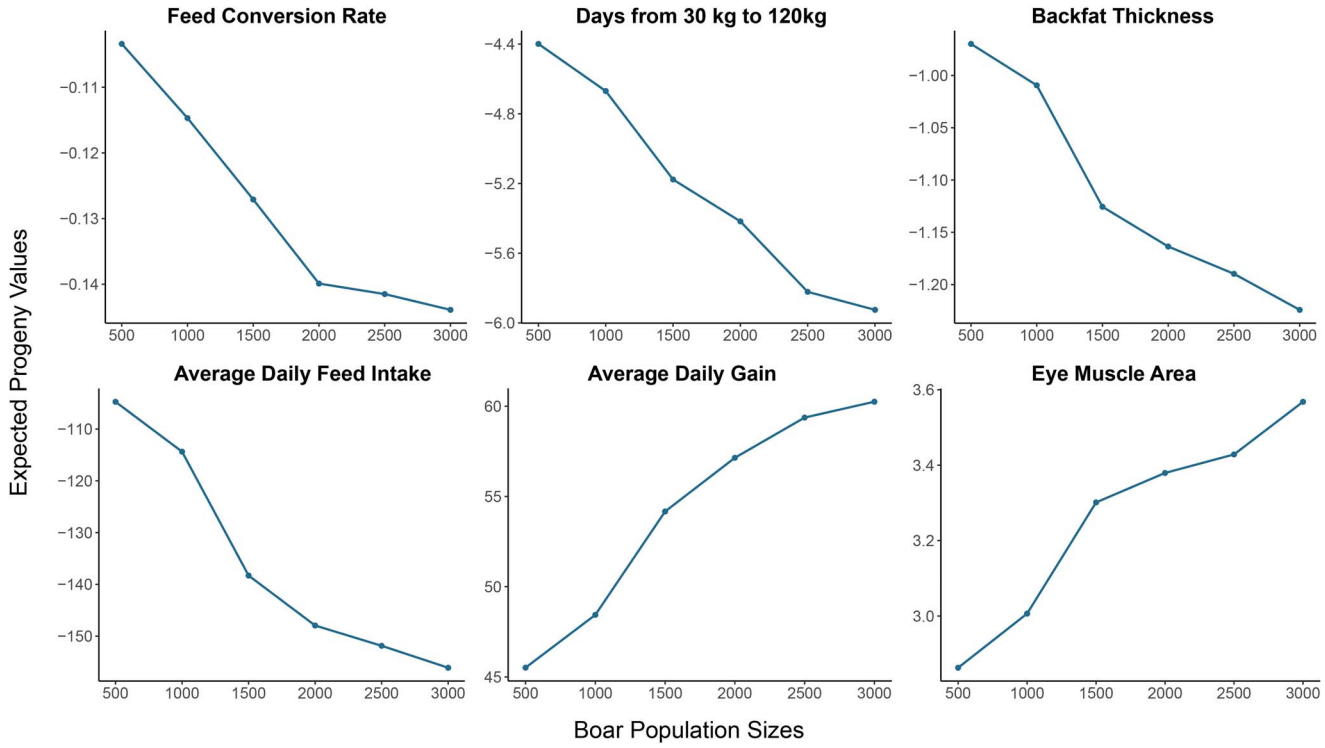

Figure 3. The performances of expected progeny values with different sizes of the boar population. The expected progeny values for six traits, which included feed conversion rate, average daily feed intake, days from 30 to 120 kg, average daily gain, backfat thickness and eye muscle area, were computed under various boar population sizes. The boar population sizes were set from 500 to 3000, and the step size was set to 500. The *X*-axis represents boar population sizes, and the *Y*-axis represents the expected progeny values of six traits.

does not require to solve MME to get the breeding values in prior, and no matter how many random effects are included in the model, it only involves the inverse of *V* matrix. Moreover, the core part *Py* can be accomplished efficiently by solving the system of linear equations on the basis of various factorization algorithms, which can completely avoid computing the direct inverse of any big matrix. Not only that, the dimension of *V* is equal to the number of effective phenotypic records, which is generally smaller than the dimension of GRM and MME that are required in traditional strategy, thus the proposed strategy should be computationally beneficial regarding the time spending and memory consumption.

Variance components estimated for a three-way crossbred population

To estimate the variance components, it is required to construct the GRM for all individuals. The method of GRM construction may affect the accuracy of estimated variance components because of the different assumptions on the distribution of genotype frequencies in a population, therefore the results would be biased when the real data structure does not match the expected hypothesis of the method used. To explore this issue, we compared four most widely used methods of GRM construction in our crossbred population, which are proposed by Zeng *et al.* [24], Su *et al.* [25], Zhu *et al.* [26] and Vitezica *et al.* [27], respectively. The results showed that there was little difference on the estimated variance components. Since the method proposed by Vitezica *et al.* has been used widely for cattle and pigs [10, 11, 28, 29], we thus used it to estimate variance components in our study. To validate our procedure, we further estimated the variance components for a published dataset using the same setting of parameters, and the results were displayed in Supplementary Table S13, which agreed with existing studies, it indicated that our working line should be correct [30, 31].

The estimated additive heritability of FCR was ∼0.29 in our study, which was higher than the report (0.20) by Miar *et al.* [32], who used pedigree information for estimation. The estimated additive heritability of AGE was 0.21, which was in agreement with the report (0.22) by Esfandyari *et al.* [33]. The additive heritability of BFT was 0.20, which was lower than the report (0.43) by Godinho *et al.* [34]. In previous studies, genetic correlation between traits in DLY three-way crossbred pigs was virtually estimated from using pedigree information. It reported that additive genetic correlation was −0.19 between FCR and ADG, 0.39 between FCR and BFT and −0.24 between BFT and EMA in commercial pigs [32, 35]. These results were slightly different with our estimates, which were largely attributed to the utilization of genomic information in this study, we therefore believe our results bring new sights and understandings of genetic architecture for complex agricultural traits.

Effects of boar population sizes on genomic mating

The larger the boar population, the greater the probability that the sow would mate with a suitable boar and, therefore, the better would be the performances of progenies. We recommend increasing the number of candidate boars to enrich the genetic source of the boar population in the application of genomic mating.

Limitations and the future of applying genomic mating

For experimental validation of genomic mating, although the practical experiment was not executed perfectly, the differences on performances of progenies between the RM group and the

GM group were significant as expected and were consistent with the theoretical simulation analyses. However, it is not that easy to fully implement the designed mating schemes, which may need to be coordinated in real practice for some of uncontrollable reasons, such as the oestrus management of sows and the semen allocation of boars.

The accurate estimation of marker effect is crucial for a successful genomic mating. It requires a continuous collection of performance measurements and genotyping data of three-way pigs, and re-evaluation of the additive and dominant effects of markers [22]. Evidence showed that assigning appropriate weights to the markers can balance its genetic contributions and to significantly improve the prediction performance of genomic selection [36], thus it is worth to try this weight strategy in genomic mating.

Conclusion

In summary, genomic mating could transform the 'breed' level of traditional synthetic line production to the 'individual' level. For the DLY three-way crossbred system, we have developed a complete genomic mating procedure with efficient algorithm for estimating marker effects for allocating the appropriate boar-sow pair to maximize progeny performances. There is no need to compute the inverse of the GRM, and there was no loss of computational accuracy. We believe that genomic mating will bring additional benefits to the current hybrid production system in pigs and in other domestic animals with the development of state-of-art algorithms for estimating marker effects and the optimization of the genomic mating procedure.

Key Points

- An 'individual level' genomic mating procedure was proposed and can be applied to commercial pig production with efficient algorithms for estimating marker effects and for allocating the appropriate boar-sow pairs, which can be freely accessed to public in our developed HIBLUP software at https://www.hiblup.com/tutorials#genomic-mating.
- Results showed that genomic mating significantly improved the performances of progeny across different traits compared with random mating, which were consistent with the real experimental validations.
- A herd of boars from a richer genetic source will increase the effectiveness of genomic mating further.

Data availability

The dataset is available at https://figshare.com/articles/dataset/Genomic_mating_using_three-way_crossbred_pigs/21362616.

Author contributions

Shuhong Zhao, Xinyun Li and Xiaolei Liu conceived and supervised the study. Xiaolei Liu, Lilin Yin and Zhenshuang Tang designed the experiments. Zhenshuang Tang conducted all the analyses with the assistance or guidance from Lilin Yin, Yunxiang Zhao, Zhiquan Wang, Xiaolei Liu, Xinyun Li and Shuhong Zhao. Lilin Yin and Haohao Zhang developed the algorithms and software. Dong Yin simulated the data. Zhenshuang Tang, Xiaolei Liu and Lilin Yin wrote the raw manuscript, Shuhong Zhao and Xinyun Li edited and proofread the manuscript. All the authors approved the final version of the manuscript.

Acknowledgments

We would like to thank Jing Xu and Liangliang Fu for their efforts in the maintenance of the experimental instruments and computational platform to guarantee our project ran smoothly.

Funding

National Natural Science Foundation of China (32072725), the Key Project of Natural Science Foundation of Hubei Province (2020CFA006) and the China Agriculture Research System of MOF and MARA (CARS-35).

References

1. Mehrabi Z, Gill M, Wijk MV, *et al.* Livestock policy for sustainable development. *Nature Food* 2020;**1**:160–5.
2. Patience JF, Rossoni-Serão MC, Gutiérrez NA. A review of feed efficiency in swine: biology and application. *J Anim Sci Biotechnol* 2015;**6**:33.
3. Tang Z, Fu Y, Xu J, *et al.* Discovery of selection-driven genetic differences of Duroc, Landrace, and Yorkshire pig breeds by EigenGWAS and Fst analyses. *Anim Genet* 2020;**51**:531–40.
4. Lin Y, Tang Q, Li Y, *et al.* Genomic analyses provide insights into breed-of-origin effects from purebreds on three-way crossbred pigs. *PeerJ* 2019;**7**:e8009.
5. Allaire FR. Mate selection by selection index theory. *Theor Appl Genet* 1980;**57**:267–72.
6. Kinghorn BP. An algorithm for efficient constrained mate selection. *Genet Sel Evol* 2011;**43**:4.
7. Destefano AL, Hoeschele I. Utilization of dominance variance through mate allocation strategies. *J Dairy Sci* 1992;**75**:1680–90.
8. Hayes BJ, Miller SP. Mate selection strategies to exploit across- and within-breed dominance variation. *J Anim Breed Genet—Zeitschrift Fur Tierzuchtung Und Zuchtungsbiologie* 2000;**117**:347–59.
9. Weigel KA, Lin SW. Use of computerized mate selection programs to control inbreeding of Holstein and Jersey cattle in the next generation. *J Dairy Sci* 2000;**83**:822–8.
10. Gonzalez-Dieguez D, Tusell L, Carillier-Jacquin C, *et al.* SNP-based mate allocation strategies to maximize total genetic value in pigs. *Genet Sel Evol* 2019;**51**:55.
11. Aliloo H, Pryce JE, Gonzalez-Recio O, *et al.* Including nonadditive genetic effects in mating programs to maximize dairy farm profitability. *J Dairy Sci* 2017;**100**:1203–22.
12. Falconer DS. *Introduction to Quantitative Genetics*. England: Addison Wesley Longman, 1996.
13. Toro MA, Varona L. A note on mate allocation for dominance handling in genomic selection. *Genet Sel Evol* 2010;**42**:33.
14. Casey DS, Stern HS, Dekkers JCM. Identification of errors and factors associated with errors in data from electronic swine feeders1. *J Anim Sci* 2005;**83**:969–82.
15. Jiao S, Maltecca C, Gray KA, *et al.* Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: I. Genetic parameter estimation and accuracy of genomic prediction. *J Anim Sci* 2014;**92**:2377–86.
16. Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. *Am J Hum Genet* 2018;**103**:338–48.
17. Chang CC, Chow CC, Tellier LC, *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;**4**:7.

18. Tang Y, Liu XL. G2P: a genome-wide-association-study simulation tool for genotype simulation, phenotype simulation and power evaluation. *Bioinformatics* 2019;**35**:3852–4.
19. Yin L, Zhang H, Tang Z, *et al.* rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genom Proteom Bioinform* 2021; **19**:619–28.
20. Vitezica ZG, Legarra A, Toro MA, *et al.* Orthogonal estimates of variances for additive, dominance, and epistatic effects in populations. *Genetics* 2017;**206**:1297–307.
21. Johnson DL, Thompson R. Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. *J Dairy Sci* 1995;**78**:449–56.
22. González-Diéguez D, Tusell L, Bouquet A, *et al.* Purebred and crossbred genomic evaluation and mate allocation strategies to exploit dominance in pig crossbreeding schemes. *G3 Genes|Genomes|Genetics* 2020;**10**:2829–41.
23. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci* 2008;**91**:4414–23.
24. Zeng Z-B, Wang T, Zou W. Modeling quantitative trait loci and interpretation of models. *Genetics* 2005;**169**:1711–25.
25. Su G, Of C, Ostersen T, *et al.* Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. *Plos One* 2012;**7**:e45293.
26. Zhu Z, Bakshi A, Vinkhuyzen AA, *et al.* Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am J Hum Genet* 2015;**96**:377–85.
27. Vitezica ZG, Varona L, Legarra A. On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics* 2013;**195**:1223–30.
28. Lopes MS, Bastiaansen JW, Janss L, *et al.* Genomic prediction of growth in pigs based on a model including additive and dominance effects. *J Anim Breed Genet* 2016;**133**: 180–6.
29. Xiang T, Christensen OF, Vitezica ZG, *et al.* Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. *Genet Sel Evol* 2016;**48**:92.
30. Zhang H, Yin L, Wang M, *et al.* Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. *Front Genet* 2019;**10**:189.
31. Cleveland MA, Hickey JM, Forni S. A common dataset for genomic analysis of livestock populations. *G3 Genes|Genomes|Genetics* 2012;**2**:429–35.
32. Miar Y, Plastow G, Bruce H, *et al.* Genetic and phenotypic correlations between performance traits with meat quality and carcass characteristics in commercial crossbred pigs. *Plos One* 2014;**9**:e110105.
33. Esfandyari H, Thekkoot D, Kemp R, *et al.* Genetic parameters and purebred–crossbred genetic correlations for growth, meat quality, and carcass traits in pigs. *J Anim Sci* 2020;**98**: 1–12.
34. Godinho RM, Bergsma R, Silva FF, *et al.* Genetic correlations between feed efficiency traits, and growth performance and carcass traits in purebred and crossbred pigs. *J Anim Sci* 2018;**96**: 817–29.
35. Miar Y, Plastow GS, Moore SS, *et al.* Genetic and phenotypic parameters for carcass and meat quality traits in commercial crossbred pigs. *J Anim Sci* 2014;**92**:2869–84.
36. Yin L, Zhang H, Zhou X, *et al.* KAML: improving genomic prediction accuracy of complex traits using machine learning determined parameters. *Genome Biol* 2020;**21**:146.