

RESEARCH ARTICLE

Open Access



Prediction of genetic merit for growth rate in pigs using animal models with indirect genetic effects and genomic information

Bjarke G. Poulsen^{1,2*} , Birgitte Ask², Hanne M. Nielsen^{1,2}, Tage Ostensen² and Ole F. Christensen¹

Abstract

Background: Several studies have found that the growth rate of a pig is influenced by the genetics of the group members (indirect genetic effects). Accounting for these indirect genetic effects in a selection program may increase genetic progress for growth rate. However, indirect genetic effects are small and difficult to predict accurately. Genomic information may increase the ability to predict indirect genetic effects. Thus, the objective of this study was to test whether including indirect genetic effects in the animal model increases the predictive performance when genetic effects are predicted with genomic relationships. In total, 11,255 pigs were phenotyped for average daily gain between 30 and 94 kg, and 10,995 of these pigs were genotyped. Two relationship matrices were used: a numerator relationship matrix (**A**) and a combined pedigree and genomic relationship matrix (**H**); and two different animal models were used: an animal model with only direct genetic effects and an animal model with both direct and indirect genetic effects. The predictive performance of the models was defined as the Pearson correlation between corrected phenotypes and predicted genetic levels. The predicted genetic level of a pig was either its direct genetic effect or the sum of its direct genetic effect and the indirect genetic effects of its group members (total genetic effect).

Results: The highest predictive performance was achieved when total genetic effects were predicted with genomic information (21.2 vs. 14.7%). In general, the predictive performance was greater for total genetic effects than for direct genetic effects (0.1 to 0.5% greater; not statistically significant). Both types of genetic effects had greater predictive performance when they were predicted with **H** rather than **A** (5.9 to 6.3%). The difference between predictive performances of total genetic effects and direct genetic effects was smaller when **H** was used rather than **A**.

Conclusions: This study provides evidence that: (1) corrected phenotypes are better predicted with total genetic effects than with direct genetic effects only; (2) both direct genetic effects and indirect genetic effects are better predicted with **H** than **A**; (3) using **H** rather than **A** primarily improves the predictive performance of direct genetic effects.

Background

Quantitative geneticists are increasingly interested in estimating genetic parameters for the impact of an individual on the phenotypes of other individuals [1]. The expectation is that selection for such beneficial interactions between animals for a certain trait will increase

genetic progress in that trait [2]. Pigs interact with one another and they have been subject to genetic analyses of interactions [3–5]. Since these interactions may give rise to previously unexploited heritable variation [6], inclusion of genetic effects of interactions between pigs in animal models for the estimation of breeding values may improve the efficiency of breeding programs [7].

Griffing [6] proposed an animal model for estimating simultaneously both an additive genetic effect on own performance (direct genetic effect) and an additive genetic effect on the performance of other individuals

*Correspondence: bgp@qgg.au.dk

¹ Center for Quantitative Genetics and Genomics, Blichers Allé 20, 8830 Tjele, Denmark

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



© The Author(s) 2020. 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.

(indirect genetic effect) on the same trait. Using these models should enable the utilization of the heritable (co) variation that arises from these interactions [7]. In addition, the association between direct and indirect genetic effects may explain why some selection schemes result in unexpected responses, whether smaller or larger than expected [6–8].

Indirect genetic effects have proven challenging to quantify and predict. Some studies on growth rate in pigs show that including indirect genetic effects in the animal model improved its goodness-of-fit and/or its predictive performance [4, 9, 10], whereas other studies found neither of these results [3, 11, 12]. Some of these studies may have been challenged by both the complex nature of interactions between pigs and the fact that indirect genetic models are more sensitive to experimental design/data structure than classical animal models [9, 13–17]. A major challenge with predicting indirect genetic effects is that they are often small and thereby require more information for accurate prediction than direct genetic effects.

Modelling indirect genetic effects with genomic information may offer a solution to the abovementioned challenge since estimating genetic relationships between pigs using genomic information rather than pedigree information increases the accuracies of the coefficients of kinship. While the estimated genetic relationship from pedigree information depends on the number of generations that is traced [18], genomic information captures both signals of selection and drift from all previous generations [19]. Consequently, genomic information is better for tracing direct genetic effects than pedigree information and could be better for indirect genetic effects too.

For direct genetic effects, estimating the genetic relationships between pigs using genomic information rather than pedigree information may result in smaller estimated variance components [20]. This may have both analytical and genomic causes [21–24]. Therefore, it is not advised to estimate variance components using genomic information [24].

In spite of the challenges associated with parameter estimation, genomic information has proven very beneficial when predicting direct genetic effects [25–27] and can be expected to be beneficial when predicting indirect genetic effects as well. This is indicated in the study by Alemu et al. [28] that predicted the sums of direct and indirect genetic effects and found that genomic information improved the predictive performance. However, their study does not answer the question of whether genomic information increased the predictive performance of direct genetic effects, indirect genetic effects or both. To our knowledge, no study has investigated whether indirect genetic effects increase the predictive

performance of animal models when direct and indirect genetic effects are predicted simultaneously with genomic information.

We hypothesize that: (1) the combined predictive performance of indirect and direct genetic effects is superior to the predictive performance of only direct effects; and (2) both prediction with the combination of indirect and direct genetic effects and prediction with direct genetic effects only is more accurate with genomic relationships than with pedigree relationships.

Methods

In this section, we successively present (1) the phenotypic data; (2) the definitions of standardized starting weight covariates; (3) the characteristics, imputation, and quality control of the genotypic data; (4) the construction of relationship matrices with either pedigree information only or both pedigree and genomic information; (5) the linear mixed models with either both indirect- and direct- genetic effects or direct genetic effects only; (6) the equations to calculate variance components and heritability statistics; and (7) the procedure used to estimate the predictive performances of genetic levels from the linear mixed models.

No Animal Care and Use Committee approval was obtained for this study because the pigs were part of routine performance tests in the DanBred breeding program. All pigs were kept in accordance with both Danish legislation for pig production and the Danish Product Standard [29].

All data was provided by SEGES, Danish Pig Research Centre.

Phenotypic data

The data was recorded in a DanBred Landrace nucleus herd in which pigs were performance-tested for average daily gain between August 2015 and October 2018. Boars and gilts were housed in different stables, and pigs were produced on-farm through artificial insemination. The test scheme was carried out on a weekly basis. The pigs entered the test when they weighed more than 28 kg with the restriction that their entry date could not be later than 2 weeks after the first pig in the pen entered the test. The test of a group ended when the largest pig in the pen reached 94 kg. The age at test start and number of days in test differed between pen-mates, since pigs within a group could enter the test at different dates. All pigs were weighed at the start and end of the test. Growth rate was defined as the average daily gain in live weight from start to end of the test; i.e. average daily gain is the average weight gain per day during the test period. Pigs that were removed from the pen before the end of the test obtained no measurement on daily gain, but were kept in

the dataset (165 pigs). The final dataset contained 11,420 pigs (6422 boars and 4833 gilts) among which 11,255 pigs had an average daily gain measurement, 1179 dams of phenotyped animals, 384 sires of phenotyped animals, 1197 groups (boar groups: 646; gilts groups: 551), and 1595 litters.

Group structure

Pigs were allocated to groups based on liveweights prior to test start: i.e. heavy pigs were grouped with other heavy pigs and vice versa, which means that relatedness between group members was not considered when grouping the animals. The group size at test start was 9 for gilts and 10 for boars. The average group size was 9.57 pigs at test start because there were more groups with boars than with gilts. Due to removals, a pig had on average 8.5 group members per day. Groups contained pigs from on average 7.7 litters (min: 4; max 10; SD: 1.2) and 6.6 sires (min: 3; max 10; SD: 1.3). The average relatedness among group members is presented as part of the results.

Definition of covariables

Two covariables related to starting weight were defined: *sWghtAge* and *WghtDev*.

sWghtAge is the expected weight of pigs at the average age pigs were weighed. The growth rate of a pig is approximately a sigmoid function of the age (developmental stage) of the pig [30]. Thus, the observed growth rate depends on the developmental stage at which it is measured. The developmental stage of a pig depends primarily on its age and previous feed intake [31]. The pigs in this study differed in developmental stages, in particular due to variation in the age at which they entered the test (Table 1). In an attempt to account for these differences in developmental stage of the pigs in the animal model, starting weights were corrected for the age at which they were observed:

$$sWghtAge_i = sWght_i - \frac{sWght_i}{sAge_i} (sAge_i - \overline{sAge}), \tag{1}$$

where *sWghtAge_i* is the weight at test start corrected for difference in age, *sWght_i* is the observed weight at test start in kg, *sAge_i* is the age at test start in days, and \overline{sAge} is the average age in days at test start across all animals.

WghtDev is the expected deviation in start weight from the group mean at the average date the group was weighed. If a pig is lighter than its group members, it is at a disadvantage when competing for resources. To calculate *WghtDev*, first we standardized starting weights to one timepoint within each group:

Table 1 Mean, standard deviation, minimum and maximum of the phenotype and covariates

Variable	Mean	SD	Min	Max
Average daily gain (g/day)	1117.8	152.7	90.9	1580.0
<i>sWghtAge</i> (kg)	29.76	2.87	16.20	41.13
<i>sWghtDev</i> (kg)	0.01	2.45	-11.94	9.53
Average number of group members in the group (heads/day)	8.5	0.6	6.2	9.0

SD, Standard deviation; *sWghtAge*, Starting weight adjusted for age at start; *sWghtDev*, Deviation in date-adjusted starting weight from group mean

$$sWghtDate_i = sWght_i - \frac{sWght_i}{sAge_i} (sDate_i - \overline{sDate}_g) \tag{2}$$

where *sAge* is as defined for *sWghtAge*; *sDate_i* is the date at which the starting weight was observed for pig *i*; and \overline{sDate}_g is the average date at which the starting weight was observed in group *g*. Lastly, *WghtDev* was defined as the difference in *WghtDate* from the group mean:

$$sWghtDev_i = sWghtDate_i - \overline{sWghtDate}_g, \tag{3}$$

where $\overline{sWghtDate}_g$ is the average *sWghtDate* within group *g*. The mean, standard deviation, minimum, and maximum of *sWghtAge* and *sWghtDev* are in Table 1.

Genotypic data and quality control

Pigs were genotyped with the NEOPEN GeneSeek® Genomic Profiler Porcine BeadChip [32]. This chip provides information on genotypes on both autosomal and sex chromosomes. The aim was to genotype all the pigs, but due to sampling errors, only 10,998 pigs (96%) were genotyped.

Genotypes from three animals that failed DanBred's parentage tests were omitted. Genotypes were regarded as valid if their GenCall-scores (GC-score) were above 60% [33]. Quality control of the genotypic data was performed per single nucleotide polymorphism (SNP) only. A SNP was used in later steps of the analysis if its call-rate was greater than 90%; its minor allele frequency was greater than 1%; its frequency of Mendelian errors was less than 0.1%; and its *p* value for the test of Hardy–Weinberg equilibrium was greater than 10⁻⁷. The Hardy–Weinberg criterion was not applied to SNPs on the sex chromosomes. After quality control of SNPs, missing genotypes were imputed with the FImpute v2.2 software [34]. In total, 34,123 SNPs and genomic information from 10,995 pigs was used for further analysis.

Relationship matrices

Genetic effects were predicted with two types of relationship matrices: a numerator relationship matrix (**A**), and a combined pedigree and genomic relationship matrix (single-step; **H**). **A** was constructed with a pedigree containing ancestors five generations prior to pigs with phenotypes. **H** was constructed from **A** and a genomic relationship matrix (**G**) using the single-step method [26, 35]. **G** was constructed using Method 1 of VanRaden [36] with centering and scaling based on the allelic frequencies observed for all genotyped pigs in the pedigree. The diagonal and off-diagonal elements of **G** were scaled and centered to have the same mean as the part of **A** that relates to genotyped animals [37–40]. Lastly, **G** was replaced by the weighted average between **G** and **A** to take into account that marker genotypes do not capture all the genetic variance:

$$\mathbf{G}_w = w\mathbf{G} + (1 - w)\mathbf{A}, \tag{4}$$

where *w* is the weight on the genomic relationship matrix. The value of *w* ranged from 0 to 99% across prediction analyses (see section on “Predictive performance”).

Animal models

Growth rate was analyzed with two linear mixed models: an animal model with both direct and indirect genetic effects (INDIRECT) and a classical animal model with only direct genetic effects (CLASSIC). For both models, variance components were estimated with **A** as covariance structure for genetic effects. Estimated variance components will be referred to by using the names of the models (DIRECT/INDIRECT). Variance component estimates and their respective model were then used to predict genetic effects with either **A** or **H** as covariance structure for genetic effects. In the following, prediction of genetic effects using CLASSIC and **A** as covariance structure will be referred to as CLASSIC_PED, predic-

Variance components were estimated with the average information restricted maximum likelihood (AI-REML) algorithm and genetic effects were predicted with best linear unbiased prediction (BLUP). Both the estimation of variance components and BLUP were carried out with the DMU software [41].

INDIRECT, the model with both indirect and direct genetic effects is as follows:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Z}_I\mathbf{a}_I + \mathbf{W}\mathbf{u} + \mathbf{Q}_D\mathbf{l}_D + \mathbf{Q}_I\mathbf{l}_I + \mathbf{R}_I\mathbf{e}_I + \mathbf{e}_D, \tag{5}$$

where subscripts D and I indicate whether an effect is direct or indirect, and the absence of a subscript indicates that the effect is expected to capture both direct and indirect effects; **y** is a vector of growth rates during the performance test; **b** is a vector of parameters for the fixed effects. The fixed class effects are sex and the interaction level between year of birth and month of birth. The fixed continuous effects are starting weight (sWghtAge; see Eq. (2)), deviation in starting weight from the group mean (sWghtDev; see Eqs. (3) and (4)), and the average number of group members present in the pen during the performance test (heads/day). The latter is used to correct for the effect of stocking density as pigs were removed during the test; **a_D** is a vector of direct genetic effects; **a_I** is a vector of indirect genetic effects; **u** is a vector of sex-year-month effects; **l_D** is a vector of direct litter effects; **l_I** is a vector of indirect litter effects, i.e. effect of litters of group mates; **e_I** is a vector of indirect environmental animal effects, i.e. this effect is analogous to a group effect; **e_D** is a vector of residuals; and **X**, **Z_D**, **Z_I**, **W**, **Q_D**, **Q_I**, and **R_I** are design matrices. Model INDIRECT contains two vectors of random genetic effects (direct and indirect) and five vectors of random, independent environmental effects (sex-year-month, direct litter, indirect litter, indirect environmental animal and residual), assumed to follow the distribution:

$$\begin{pmatrix} \mathbf{a}_D \\ \mathbf{a}_I \\ \mathbf{u} \\ \mathbf{l}_D \\ \mathbf{l}_I \\ \mathbf{e}_I \\ \mathbf{e}_D \end{pmatrix} \sim \text{MVN} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_{a_D}^2 & \sigma_{a_{D,I}} \\ \sigma_{a_{D,I}} & \sigma_{a_I}^2 \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \otimes \mathbf{C} \begin{pmatrix} \sigma_u^2 \otimes \mathbf{I}_u & & & & & & \\ & \sigma_{l_D}^2 \otimes \mathbf{I}_{l_D} & & & & & \\ & & \sigma_{l_I}^2 \otimes \mathbf{I}_{l_I} & & & & \\ & & & \sigma_{e_I}^2 \otimes \mathbf{I}_{e_I} & & & \\ & & & & \sigma_{e_D}^2 \otimes \mathbf{I}_{e_D} & & \end{pmatrix} \right), \tag{6}$$

tion of genetic effects using CLASSIC and **H** as covariance structure will be referred to as CLASSIC_GEN, prediction of genetic effects using INDIRECT and **A** as covariance structure will be referred to as INDIRECT_PED, and prediction of genetic effects using INDIRECT and **H** as covariance structure will be referred to as INDIRECT_GEN.

where $\sigma_{a_D}^2$ is the direct genetic variance; $\sigma_{a_I}^2$ is the indirect genetic variance; $\sigma_{a_{D,I}}$ is the covariance between direct and indirect genetic effects; **C** is the covariance structure for genetic effects, i.e. **C** is **A** for models INDIRECT and INDIRECT_PED, and **H** for model INDIRECT_GEN; σ_u^2 is the sex-year-month variance; $\sigma_{l_D}^2$ is

the direct litter variance; $\sigma_{l_1}^2$ is the indirect litter variance; $\sigma_{e_1}^2$ is the indirect non-genetic animal variance; $\sigma_{e_D}^2$ is the residual variance; and \mathbf{I}_u , \mathbf{I}_{l_D} , \mathbf{I}_{l_1} , \mathbf{I}_{e_1} , and \mathbf{I}_{e_D} are identity matrices with dimensions equal to the number of levels of their respective effects.

CLASSIC, the model with only direct genetic effects, is as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Wu} + \mathbf{Q}_D\mathbf{l}_D + \mathbf{Q}_l\mathbf{l}_1 + \mathbf{R}_1\mathbf{e}_1 + \mathbf{e}_D, \tag{7}$$

where matrix and vector notations are identical to those for INDIRECT (Eq. (5)). CLASSIC contains one vector of random genetic effects (direct) and five vectors of random, independent environmental effects (sex-year-month, direct litter, indirect litter, indirect environmental animal and residual), assumed to follow the distribution:

$$\begin{bmatrix} \mathbf{a}_D \\ \mathbf{u} \\ \mathbf{l}_D \\ \mathbf{l}_1 \\ \mathbf{e}_1 \\ \mathbf{e}_D \end{bmatrix} \sim \text{MVN} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_{a_D}^2 \otimes \mathbf{C} & & & & & \\ \mathbf{0} & \sigma_u^2 \otimes \mathbf{I}_u & & & & \\ \mathbf{0} & \mathbf{0} & \sigma_{l_D}^2 \otimes \mathbf{I}_{l_D} & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{l_1}^2 \otimes \mathbf{I}_{l_1} & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{e_1}^2 \otimes \mathbf{I}_{e_1} & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{e_D}^2 \otimes \mathbf{I}_{e_D} \end{bmatrix} \right), \tag{8}$$

where the notation is the same as for INDIRECT (Eq. (6)).

CLASSIC contains indirect litter and indirect non-genetic animal effects although it is meant to represent a usual animal model. The reasons are that these indirect effects explain phenotypic variation and that greater similarity between CLASSIC and INDIRECT makes the results on their log-likelihood-ratio test more related to whether indirect genetic effects exist or not.

A log-likelihood ratio test was used to test whether INDIRECT had a better model fit than CLASSIC: $P(\Delta \text{Log}L) = \chi^2 \left(-2 \log_e \left[\frac{\ell(\text{CLASSIC})}{\ell(\text{INDIRECT})} \right] \right)$, where $\chi^2(\dots)$ is the cumulative distribution function of the Chi square distribution with two degrees of freedom for the parameters $\sigma_{a_1}^2$ and $\sigma_{a_{D,1}}$; $\ell(\text{CLASSIC})$ is the likelihood value of CLASSIC; and $\ell(\text{INDIRECT})$ is the likelihood value of INDIRECT.

Variance components and heritability parameters

The equations used to calculate variance components and heritability statistics for both CLASSIC and INDIRECT are in Table 2. Two heritability statistics were calculated: direct heritability (h^2) and total heritability (T^2) [2].

Predictive performance

To estimate the predictive performance of the models, the data were divided into training data and validation data. Training data contained phenotypes on pigs in

Table 2 Definitions of variance components and heritability statistics

	Classic	Indirect
σ_{TBV}^2	None	$\sigma_{a_D}^2 + 2(\bar{n} - 1)\sigma_{a_{D,l}} + (\bar{n} - 1)^2\sigma_{a_l}^2$
σ_{env}^2	$\sigma_u^2 + \sigma_{l_D}^2 + (\bar{n} - 1)(\sigma_{l_1}^2 + \sigma_{e_1}^2) + \sigma_{e_D}^2$	
σ_p^2	$\sigma_{a_D}^2 + \sigma_{env}^2$	$\sigma_{a_D}^2 + (\bar{n} - 1)\sigma_{a_l}^2 + r(\bar{n} - 1)(2\sigma_{a_{D,l}} + (\bar{n} - 2)\sigma_{a_l}^2) + \sigma_{env}^2$
h^2	$\sigma_{a_D}^2 * [\sigma_p^2]^{-1}$	
T^2	None	$\sigma_{TBV}^2 * [\sigma_p^2]^{-1}$

σ_{TBV}^2 , total heritable variance. $\sigma_{a_{D,l}}^2$, direct genetic variance. \bar{n} , average group size at test start (9.57 pigs). $\sigma_{a_{D,l}}$, direct-indirect genetic covariance. $\sigma_{a_l}^2$, indirect genetic variance. σ_u^2 , sex-year-month variance. $\sigma_{l_D}^2$, direct litter variance. $\sigma_{l_1}^2$, indirect litter variance. $\sigma_{e_1}^2$, Indirect animal variance. $\sigma_{e_D}^2$, residual variance. r , average relatedness among animals in a pen, averaged across pens (16.6%). σ_{env}^2 , total variance due to not-heritable variance components. σ_p^2 : total phenotypic variance. h^2 , direct heritability. T^2 , total heritability

groups where all pigs were born before January 1st, 2018 (9431 pigs; 83.8%), whereas the validation data contained phenotypes on pigs not included in the training data (1824 pigs; 16.2%).

The predictive performance of the animal models was defined as the ability to use the training data to predict corrected phenotypes of pigs in the validation data. The predictive performance was quantified as Pearson's correlation coefficient between corrected phenotypes and predicted genetic levels. Here, the corrected phenotype was defined as the difference between observed phenotypes and fixed effects from INDIRECT: $\mathbf{y}_c = \mathbf{y} - \mathbf{Xb}$. The predicted genetic level was either based on direct genetic effects only, $\mathbf{DGE}_i = \mathbf{a}_i^D$, or on a combination of direct and indirect genetic effects, $\mathbf{TGE}_i = \mathbf{a}_i^D + \sum_{j \neq i} \mathbf{a}_j^I$ (total genetic effect), where \mathbf{a}_i^D is the direct genetic effect of pig i , and \mathbf{a}_j^I is the indirect genetic effect of group member j of pig i . Predictive performance of indirect genetic effects alone was not assessed, since this is not informative for investigating whether indirect genetic effects improve the total predictive ability. A Hotelling-Williams t-test [42, 43] was used to determine whether differences in predictive performances of genetic levels were statistically significant.

For both INDIRECT_GEN and CLASSIC_GEN, genetic effects were predicted 20 times with different weights on \mathbf{G} when constructing \mathbf{H} (Eq. (4)). The weight on \mathbf{G} ranged from 5 to 99% with an interval of 5% except between $w = 95\%$ and $w = 99\%$; i.e.

$w \in \{5\%, 10\%, 15\%, \dots, 95\%, 99\%\}$ (Eq. (1)). The purpose was twofold: (1) to find the weight on **G** that returned the best predictive performance and (2) to examine the ranking between predictive performances of **DGE** and **TGE** at different weights on **G**. For both **INDIRECT_GEN** and **CLASSIC_GEN**, the genomic analysis with the best predictive performance was used to represent the predictive performance of that model.

The prediction biases of genetic levels were defined as the linear regression coefficient of corrected phenotypes on genetic levels. Just as the predictive performance, the prediction bias was calculated with different weights on **G**.

Results

Estimated variance components

Direct genetic variances were similar between **CLASSIC** and **INDIRECT** (σ_{aD}^2 : 3339 ± 450 vs. 3349 ± 451 , Table 3). The genetic correlation between direct and indirect genetic effects was not statistically significantly different from zero ($\rho_{aD,I}$: 0.06 ± 0.18). The indirect litter and indirect animal variances were larger for **CLASSIC** than for **INDIRECT** (σ_{l1}^2 : 37.7 ± 11.1 vs. 19.1 ± 12.2 ; σ_{e1}^2 : 112.9 ± 18.1 vs. 97.4 ± 21.8). The direct heritability was 0.19 ± 0.02 for both **CLASSIC** and **INDIRECT**. For **INDIRECT**, the total heritability was larger than the direct heritability (0.35 ± 0.09 vs. 0.19 ± 0.02).

Table 3 Estimates of variance components, heritability, and likelihood ratio statistics with associated p-values

	Classic	Indirect
σ_{aD}^2	3339 (± 450)	3349 (± 451)
σ_{aI}^2		34.4 (± 16.3)
$\rho_{aD,I}$		0.06 (± 0.18)
σ_u^2	951 (± 319)	908 (± 306)
σ_{lD}^2	636 (± 131)	633 (± 131)
σ_{eD}^2	10,971 (± 289)	10,989 (± 295)
σ_{lI}^2	37.7 (± 11.1)	19.1 (± 12.2)
σ_{eI}^2	112.9 (± 18.1)	97.4 (± 21.8)
σ_p^2	17,188 (± 420)	17,597 (± 495)
σ_{TBV}^2		6201 (± 1694)
h^2	0.19 (± 0.02)	0.19 (± 0.02)
T^2		0.35 (± 0.09)
$-2\text{Log}L$	119,074.46	119,067.12
$P(\Delta\text{Log}L)$		0.025

σ_{aD}^2 , direct genetic variance. σ_{aI}^2 , indirect genetic variance. $\rho_{aD,I}$, genetic correlation between direct- and indirect- genetic effects. σ_u^2 , sex-year-month variance. σ_{lD}^2 , litter variance. σ_{eD}^2 , residual variance. σ_{lI}^2 , indirect litter variance. σ_{eI}^2 , indirect animal variance. σ_p^2 , total phenotypic variance. σ_{TBV}^2 , total heritable variance. h^2 , direct heritability. T^2 , total heritability. $-2\text{Log}L$, -2 times the log-likelihood. $P(\Delta\text{Log}L)$, Chi square p-value (2 degrees of freedom) of log-likelihood ratio between **INDIRECT** and **CLASSIC**

Table 4 Predictive performances of direct genetic effects (DGE) and total genetic effects (TGE)

	Predictive performance	
	DGE	TGE
CLASSIC_PED	14.7% ^a	
CLASSIC_GEN	20.9% ^b	
INDIRECT_PED	14.7% ^a	15.3% ^a
INDIRECT_GEN	21.0% ^c	21.2% ^{bc}

Predictive performances with different superscripted letters a, b and c were statistically different according to the Hotelling-Williams t-test

Predictive performance

The predictive performance of direct genetic effect (DGE) from **CLASSIC**, DGE from **INDIRECT**, and total genetic effects (TGE) from **INDIRECT** were all improved when predicting genetic effects with the single-step relationship matrix rather than the numerator relationship matrix (Table 4 and Fig. 1). The best predictive performance was achieved with TGE from **INDIRECT** and a 50% weight on the genomic relationship matrix (Fig. 1). The predictive performance of TGE was better than the predictive performance of DGE (Table 4 and Fig. 1), although differences in predictive performance between TGE and DGE were not statistically significant (Table 4). The difference in predictive performance between TGE and DGE decreased as the weight on genomic information increased (Fig. 1; Eq. (1)).

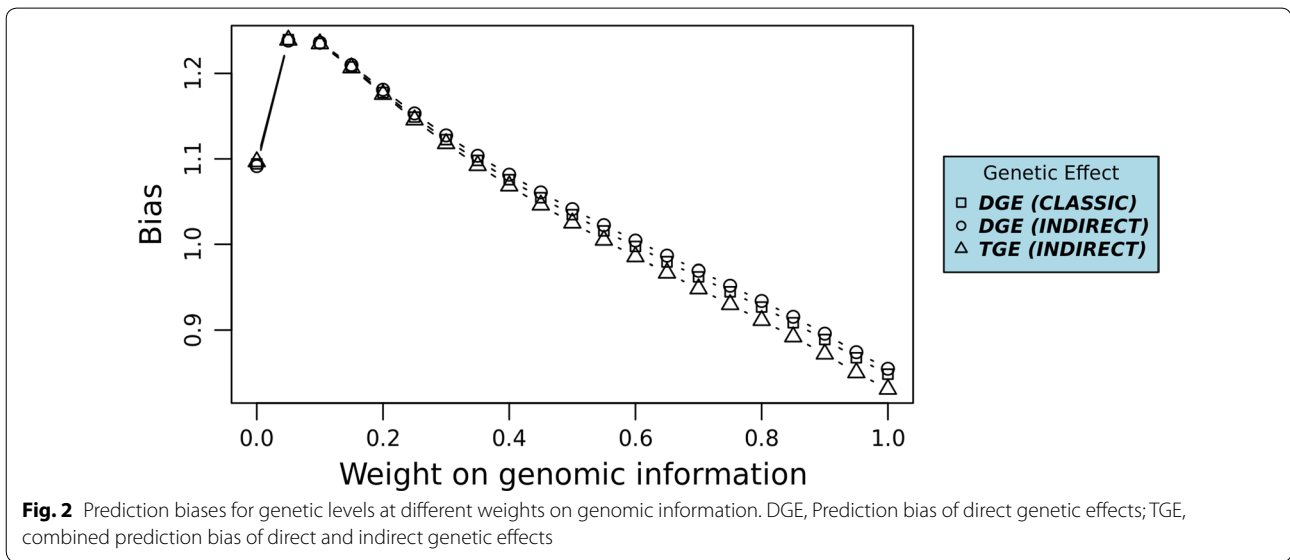
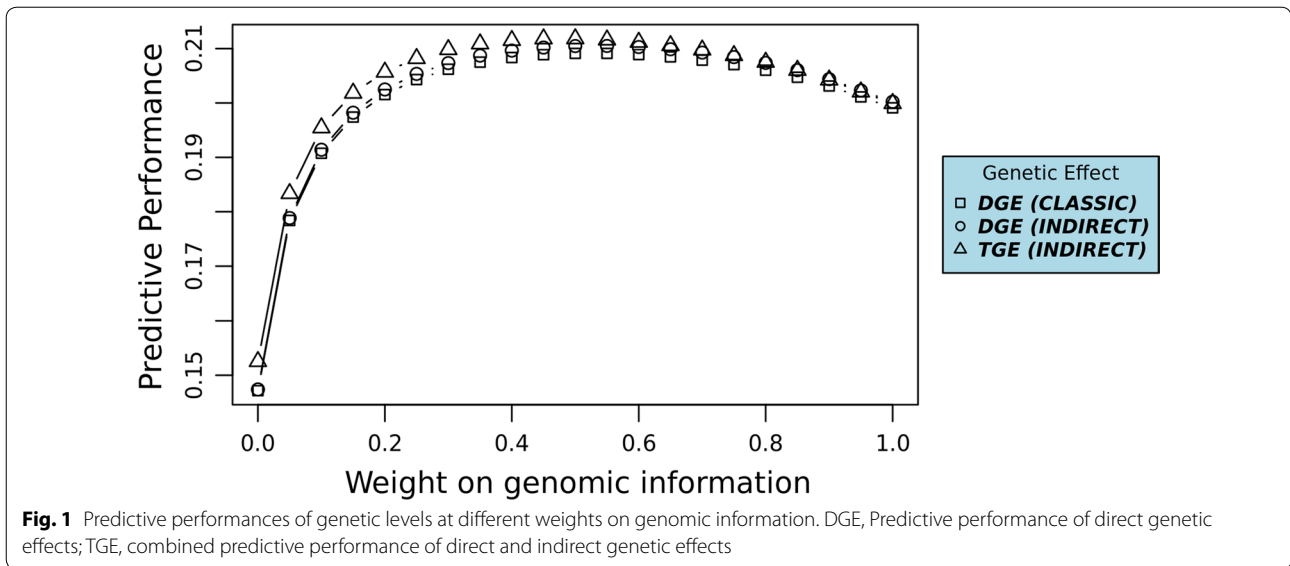
Prediction bias

The patterns and levels of prediction biases were similar for DGE and TGE (Fig. 2). Prediction bias was lowest with 45% weight on genomic information for DGE and with 50% weight on genomic information for TGE.

Discussion

As expected, the predictive performances of the models with genomic information were superior to those using pedigree information only (Table 4 and Fig. 1). Thereby, the predictive performance of the indirect genetic model was improved with genomic information. This is in accordance with the previous studies on genomic prediction of TGE by Alemu, et al. [28] and Brinker, et al. [44].

As expected, the predictive performance of TGE was better than the predictive performance of DGE for indirect genetic models. However, the predictive superiority of TGE compared to DGE decreased as the weight on genomic information increased. Although it is difficult to interpret patterns of differences in correlations, this result indicates that indirect genetic effects estimated with genomic information contribute less than indirect effects estimated with pedigree information to the predictive performance of TGE. Therefore, one may speculate that



pedigree information explains indirect genetic effects more accurately than genomic information.

Corrected phenotypes were defined as the difference between the raw phenotypes and the expected levels based on fixed effects from INDIRECT. This gives INDIRECT models an advantage over DIRECT models in the prediction analysis. Furthermore, corrected phenotypes can also be calculated as the difference between raw phenotypes and expected levels based on both fixed effects and selected random effects. The correction of phenotypes for random effects relies on the assumption of independence between the random effect used for correction and the genetic effects. If this assumption is violated, the

choice of method for calculation of corrected phenotypes can affect the ranking of predictive performances of genetic levels. In this study, we found that the method for calculating the corrected phenotypes influenced the results on predictive performance; i.e. the predictive performance of TGE was inferior to the predictive performance of DGE when correcting the phenotype for both fixed effects and random sex-year-month (results not presented). This susceptibility to method for correction of phenotypes indicates that sex-year-month effects and indirect genetic effects were not independently distributed as assumed by the animal model.

In this study, we attempted to correct for indirect fixed effects. For example, the deviation in weight at test start from the average starting weight of the group (sWghtDev; see "Methods") is both a direct and indirect covariate. Since sWghtDev was positively associated with later growth rate (results not shown), its direct interpretation is that pigs that are larger than their group members at test start grow faster themselves. However, the sum of the deviations from the mean is zero, and therefore $w_i - \bar{w} = -\left(\sum_{j \neq i} w_j - \bar{w}\right)$; i.e. the effect of sWghtDev is equivalent to the effect of the sum of the group members' sWghtDev. Thereby, the indirect interpretation of sWghtDev is that pigs that weigh less than the average value for the group make their group members grow faster.

It is important to account for both direct and indirect fixed effects when predicting indirect genetic effects [16]. An indirect genetic model can be perceived as a composite animal model that comprises a classical direct animal model component and an indirect animal model component, where each component contains fixed effects, random environmental effects, and random genetic effects. Many previous studies on indirect genetic effects sparsely correct for indirect fixed effects. This may be partly due to very strong confounding between levels of effects for the direct and indirect components in some cases, e.g. in our study, year of birth was strongly confounded with year of birth of group mates. However, indirect fixed effects and direct fixed effects are not necessarily confounded and including indirect fixed effects in the animal model may lead to improved statistical modelling and predictive performance of indirect genetic models.

Similarly to the correction for indirect fixed effects, correction for indirect random effects may affect the predictive performance of animal models. In this study, the only indirect environmental random effect was an indirect environmental animal effect; i.e. an indirect analogue of the direct residual. Most previous studies on indirect genetic effects on the growth rate of finishers have included a random group effect [3, 5, 9, 45], which is equivalent to the indirect environmental animal effect in this study when group sizes are constant. Other studies have examined other indirect effects in addition to the random group effect. For example, Canario et al. [10] and Canario et al. [46] examined the indirect effect of litter, and Nielsen et al. [9] examined the indirect effect of sex. For Canario et al. [46], including indirect litter effects in the animal model reduced the estimated indirect genetic variance by two-thirds. This could indicate that indirect genetic effects capture indirect environmental effects if these are unaccounted for in the animal model. Thus, indirect genetic variances estimated by Arango et al. [3], Duijvesteijn et al. [5], Nielsen et al. [9], Chen et al. [45]

may be overestimated due to insufficient correction for indirect environmental effects other than the group effects.

The animal models of this study corrected for both indirect environmental animal effects and indirect litter effects. Nevertheless, the estimated indirect genetic variance component from this study was larger than most previously published indirect genetic variances for finisher growth (34.4 ± 16.3 vs 6–20) [3, 5, 9, 10, 45]. This may be because our dataset only contains phenotypes from a single herd and thereby a narrower distribution of environmental effects than other studies. This is supported by the fact that we obtained larger estimates of direct genetic variances than Nielsen et al. [9], who analyzed finisher growth across more herds in the same population.

Most of the estimated variance components are of similar magnitude to those from other studies. For example, the direct genetic variance (3349 ± 451) is marginally greater than those published by other studies (2762, 2521, and 3200) [9, 47–49]; the direct litter variance (633 ± 131) is in the range of those published by other studies (984, 652, and 1576) [9, 47, 48]; and the residual variance ($10,989 \pm 295$) was greater than those published by other studies (7246, 5439, and 10,000) [9, 47–49]. We may have obtained greater estimates of residual variances because of how we modelled contemporary groups (HYM). Other studies included contemporary groups as fixed effects, while we included contemporary group as a random effect.

In our study, we assumed that indirect genetic effects were not affected by the number of pigs in the group, and not affected by whether pigs were removed from the pen prior to the end of its testing period. It has been suggested that the indirect genetic effect of a pig should be scaled (diluted) based on the number of pigs in its group or the number of days it was in the pen [50, 51]. In our study, dilution according to group size was not relevant since there was only little variation in group size (9 vs. 10). Dilution of indirect genetic effects according to the time the pig spends in the pen was investigated; however, in accordance with Ask et al. [47], dilution did not provide an increase in predictive performance, and was therefore omitted.

In our study, the effects of sex, group size, and stable are confounded. Consequently, INDIRECT assumes that all genetic (co)variance components were equal across sex to avoid confusing the effect of sex with genome by environment interactions when interpreting the results. However, Nielsen et al. [9] found that the genetic correlation between indirect genetic effects of boars and gilts was less than 1 and that both direct and indirect genetic variances differ between boars and gilts. We could not

test this in our study due to confounding of sex with stable. In future studies, it would be preferable to use data and animal models that enable the estimation of genetic variance components both within and across boars and gilts.

As mentioned above, based on the results on predictive performance it can be speculated whether indirect genetic effects could be more appropriately modelled by the numerator relationship matrix than the single-step relationship matrix. Thus, we examined an alternative to INDIRECT. In this animal model, direct and indirect genetic effects were uncorrelated; both direct and indirect genetic effects were modelled using single-step relationship matrices; and the weight on genomic information in the single-step matrix ranged from 0 to 99% for indirect genetic effects, but was fixed at 1% for direct genetic effects. The predictive performance of TGE from this alternative model was, surprisingly, superior to any other model developed during our study when indirect genetic effects were modelled with pedigree information only. Meanwhile, the predictive performance of DGE from this model did not depend on the weight on genomic information for indirect genetic effects. We do not recommend the use of this model for prediction as we can not explain this pattern and because the direct genetic effects no longer provide information on indirect genetic effects through their covariance structure. Nevertheless, the pattern of prediction of this alternative model may further indicate challenges with prediction of indirect genetic effects with genomic information. Further research is needed to investigate whether results are replicable both within other populations and with other data to assess the consistency of the results.

Our results both support and challenge prediction of genetic effects with indirect genetic models and genomic information. On one hand, both TGE and DGE from the indirect genetic models with genomic information were superior to DGE from the classical animal models with genomic information. On the other hand, the conclusions are affected by both how heavily genomic information is weighted and the choice of method for calculating corrected phenotypes. We believe that especially the latter indicates insufficient correction for both fixed and random environmental effects. The animal models of this study corrected for environmental effects similarly to previous studies [5, 9, 52]. Therefore, we are not sure that indirect genetic models with genomic informations are ready for implementation in an industrial genetic evaluation system for pigs. Instead, we strongly encourage future studies to attempt to replicate these results in other populations, and to expand the search for indirect environmental effects beyond those that have previously been associated with the trait in a classical genetic model with only direct genetic effects.

Conclusions

This study provides evidence that: (1) the combined predictive performance of indirect genetic effects and direct genetic effects is superior to the predictive performance of only direct effects; and (2) prediction of genetic levels is more accurate with genomic relationships than with pedigree relationships. The better predictive performance of models with genomic relationships is primarily due to better predictive performance of direct genetic effects. However, it is important to note that the first conclusion was sensitive towards both the definition of corrected phenotypes and the relative weight on genomic information in the single-step relationship matrix.

Acknowledgements

Thanks to Per Madsen from Center for Quantitative Genetics and Genomics - Aarhus University for assisting with the technical aspect of model development.

Authors' contributions

BGP performed all analyses and wrote the paper. The experiment was designed by BAS, HMN and TAO. BA and HMN oversaw collection of phenotypes and TO oversaw collection of genotypes. OFC and BGP developed the animal models. BA, OFC, TO and HMN provided input on the data editing procedures, data analyses, and commented on and improved drafts of the paper. All authors read and approved the final manuscript.

Funding

This study was part of the project "Breeding for feed efficiency and behavior of pigs in groups" funded by GUDP (Green Development and Demonstration Programme), Ministry of Environment and Food of Denmark, Grant No. 34009-14-0849.

Availability of data and materials

The data analyzed during this study is not publicly available as it is owned by SEGES P/S.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Center for Quantitative Genetics and Genomics, Blichers Allé 20, 8830 Tjele, Denmark. ² SEGES, Danish Pig Research Centre, Danish Agriculture and Food Council F.m.b.A., Axeltorv 3, 1609 Copenhagen V, Denmark.

Received: 29 January 2020 Accepted: 21 September 2020
Published online: 07 October 2020

References

- Ellen ED, Rodenburg TB, Albers GA, Bolhuis JE, Camerlink I, Duijvesteijn N, et al. The prospects of selection for social genetic effects to improve welfare and productivity in livestock. *Front Genet.* 2014;5:377.
- Bijma P, Muir WA, Van Arendonk JAM. Multilevel selection 1: quantitative genetics of inheritance and response to selection. *Genetics.* 2007;175:277–88.
- Arango J, Misztal I, Tsuruta S, Culbertson M, Herring W. Estimation of variance components including competitive effects of large white growing gilts. *J Anim Sci.* 2005;83:1241–6.

4. Bouwman AC, Bergsma R, Duijvesteijn N, Bijma P. Maternal and social genetic effects on average daily gain of piglets from birth until weaning. *J Anim Sci*. 2010;88:2883–92.
5. Duijvesteijn N, Knol EF, Bijma P. Estimation and validation of indirect genetic effects for average daily gain in two purebred sire lines. In Proceedings of the 10th World Congress on Genetics Applied to Livestock Production: 17–22 August 2014; Vancouver; 2014. p. 564.
6. Griffing B. Selection in reference to biological groups. I. Individual and group selection applied to populations of unordered groups. *Aust J Biol Sci*. 1967;20:127–39.
7. Bijma P, Wade MJ. The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. *J Evol Biol*. 2008;21:1175–88.
8. Muir WM. Incorporation of competitive effects in forest tree or animal breeding programs. *Genetics*. 2005;170:1247–59.
9. Nielsen HM, Ask B, Madsen P. Social genetic effects for growth in pigs differ between boars and gilts. *Genet Sel Evol*. 2018;50:4.
10. Canario L, Lundeheim N, Bijma P. Pig growth is affected by social genetic effects and social litter effects that depend on group size. In Proceedings of the 9th World Congress on Genetics Applied to Livestock Production: 1–6 August 2010; Leipzig; 2010. p. 87.
11. Camerlink I, Bergsma R, Duijvesteijn N, Bolhuis J, Bijma P. Consequences of selection for social genetic effects on adg in finishing pigs—a pilot study. In Proceedings of the 9th World Congress on Genetics Applied to Livestock Production: 1–6 August 2010; Leipzig; 2010. p. 595.
12. Hsu WL, Johnson RK, Van Vleck LD. Effect of pen mates on growth, backfat depth, and longissimus muscle area of swine. *J Anim Sci*. 2010;88:895–902.
13. Hamilton WD. The genetical evolution of social behaviour. I. *J Theor Biol*. 1964;7:1–16.
14. Komdeur J, Hatchwell BJ. Kin recognition: function and mechanism in avian societies. *Trends Ecol Evol*. 1999;14:237–41.
15. Bijma P, Muir WM, Ellen ED, Wolf JB, Van Arendonk JA. Multilevel selection 2: estimating the genetic parameters determining inheritance and response to selection. *Genetics*. 2007;175:289–99.
16. Bijma P. The quantitative genetics of indirect genetic effects: a selective review of modelling issues. *Heredity* (Edinb). 2014;112:61–9.
17. Bijma P. Estimating indirect genetic effects: precision of estimates and optimum designs. *Genetics*. 2010;186:1013–28.
18. Mrode RA. Linear models for the prediction of animal breeding values. 3rd ed. Wallingford: CAB; 2014.
19. Nielsen R, Slatkin M. An introduction to population genetics: theory and applications. 1st ed. Sunderland: Sinauer Associates Inc.; 2013.
20. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of snp-based heritability. *Nat Genet*. 2017;49:1304–10.
21. Zhang C, Kemp RA, Stothard P, Wang Z, Boddicker N, Krivushin K, et al. Genomic evaluation of feed efficiency component traits in duroc pigs using 80 k, 650 k and whole-genome sequence variants. *Genet Sel Evol*. 2018;50:14.
22. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc Natl Acad Sci USA*. 2012;109:1193–8.
23. Meuwissen T, Goddard M. Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics*. 2010;185:623–31.
24. de los Campos G, Sorensen D, Gianola D. Genomic heritability: What is it? *PLoS Genet*. 2015;11:e1005048.
25. Knol EF, Nielsen B, Knap PW. Genomic selection in commercial pig breeding. *Anim Front*. 2016;6:15–22.
26. Christensen OF, Lund MS. Genomic prediction when some animals are not genotyped. *Genet Sel Evol*. 2010;42:2.
27. Hidalgo AM, Bastiaansen JWM, Lopes MS, Harlizius B, Groenen MAM, de Koning DJ. Accuracy of predicted genomic breeding values in purebred and crossbred pigs. G3 (Bethesda). 2015;5:1575–83.
28. Alemu SW, Calus MPL, Muir WM, Peeters K, Vereijken A, Bijma P. Genomic prediction of survival time in a population of brown laying hens showing cannibalistic behavior. *Genet Sel Evol*. 2016;48:68.
29. SEGES. <https://pigresearchcentre.dk/DANISH-Standards/-/media/38F57DF2D3884B88AE8C79178E2F1DCE.ashx>. Accessed 17 Jun 2019.
30. Fitzhugh HA Jr. Analysis of growth curves and strategies for altering their shape. *J Anim Sci*. 1976;42:1036–51.
31. Le Cozier Y, Ringmar-Cederberg E, Johansen S, Dourmad JY, Neil M, Sterna S. Effect of feeding level during rearing and mating strategy on performance of swedish yorkshire sows. 1. Growth, puberty and conception rate. *Anim Sci*. 1999;68:355–63.
32. Neogen GeneSeek Operations. https://genomics.neogen.com/pdf/slicks/ag284_ggp_porcine.pdf. Accessed 02 Jul 2019.
33. Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. Beadarray™ technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques*. 2002;32:56–61.
34. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*. 2014;15:478.
35. Legarra A, Christensen OF, Aguilar I, Misztal I. Single step, a general approach for genomic selection. *Livest Sci*. 2014;166:54–65.
36. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91:4414–23.
37. Chen CY, Misztal I, Aguilar I, Legarra A, Muir WM. Effect of different genomic relationship matrices on accuracy and scale. *J Anim Sci*. 2011;89:2673–9.
38. Christensen OF, Madsen P, Nielsen B, Ostersen T, Su G. Single-step methods for genomic evaluation in pigs. *Animal*. 2012;6:1565–71.
39. Vitezica ZG, Aguilar I, Misztal I, Legarra A. Bias in genomic predictions for populations under selection. *Genet Res (Camb)*. 2011;93:357–66.
40. Forni S, Aguilar I, Misztal I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet Sel Evol*. 2011;43:1.
41. Madsen P, Jensen J. A user's guide to DMU: a package for analysing multivariate mixed models. Version 6, release 5.2. Tjele: Center for Quantitative Genetics and Genomics; 2013.
42. Dunn OJ, Clark V. Comparison of tests of the equality of dependent correlation coefficients. *J Am Stat Assoc*. 1971;66:904–8.
43. Revelle W. *Psych: Procedures for psychological, psychometric, and personality research*. Evanston: Northwestern University; 2018.
44. Brinker T, Raymond B, Bijma P, Vereijken A, Ellen ED. Estimation of total genetic effects for survival time in crossbred laying hens showing cannibalism, using pedigree or genomic information. *J Anim Breed Genet*. 2017;134:60–8.
45. Chen CY, Kachman SD, Johnson RK, Newman S, Van Vleck LD. Estimation of genetic parameters for average daily gain using models with competition effects. *J Anim Sci*. 2008;86:2525–30.
46. Canario L, Lundeheim N, Bijma P. The early-life environment of a pig shapes the phenotypes of its social partners in adulthood. *Heredity* (Edinb). 2017;118:534–41.
47. Ask B, Christensen OF, Heidaritabar M, Madsen P, Nielsen HM. The predictive ability of indirect genetic models is reduced when culled animals are omitted from the data. *Genet Sel Evol*. 2020;52:8.
48. Kavlak AT, Uimari P. Estimation of heritability of feeding behaviour traits and their correlation with production traits in finnish yorkshire pigs. *J Anim Breed Genet*. 2019;136:484–94.
49. Do DN, Strathe AB, Jensen J, Mark T, Kadarmideen HN. Genetic parameters for different measures of feed efficiency and related traits in boars of three pig breeds. *J Anim Sci*. 2013;91:4069–79.
50. Bijma P. Multilevel selection 4: modeling the relationship of indirect genetic effects and group size. *Genetics*. 2010;186:1029–31.
51. Cappa EP, Cantet RJ. Direct and competition additive effects in tree breeding: bayesian estimation from an individual tree mixed model. *Silvae Genet*. 2008;57:45–56.
52. Bergsma R, Kanis E, Knol EF, Bijma P. The contribution of social effects to heritable variation in finishing traits of domestic pigs (*Sus scrofa*). *Genetics*. 2008;178:1559–70.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.