Journal of Animal Science, 2020, 1–10

doi:10.1093/jas/skaa034
Advance Access publication February 04, 2020
Received: 16 August 2019 and **Accepted:** 10 February 2020
Animal Genetics and Genomics

ANIMAL GENETICS AND GENOMICS

Optimizing design to estimate genetic correlations between environments with common environmental effects

Maria Lozano-Jaramillo,[1] Hans Komen, Yvonne C. J. Wientjes, Han A. Mulder, and John W. M. Bastiaansen

Animal Breeding and Genomics Group, Wageningen University & Research, Wageningen, The Netherlands

[1]Corresponding author: m.lozanojaramillo@gmail.com

ORCiD number: 0000-0001-8344-4827 (M. Lozano-Jaramillo).

Abstract

Breeding programs for different species aim to improve performance by testing members of full-sib (**FS**) and half-sib (**HS**) families in different environments. When genotypes respond differently to changes in the environment, this is defined as genotype by environment (G × E) interaction. The presence of common environmental effects within families generates covariance between siblings, and these effects should be taken into account when estimating a genetic correlation. Therefore, an optimal design should be established to accurately estimate the genetic correlation between environments in the presence of common environmental effects. We used stochastic simulation to find the optimal population structure using a combination of FS and HS groups with different levels of common environmental effects. Results show that in a population with a constant population size of 2,000 individuals per environment, ignoring common environmental effects when they are present in the population will lead to an upward bias in the estimated genetic correlation of on average 0.3 when the true genetic correlation is 0.5. When no common environmental effects are present in the population, the lowest standard error (**SE**) of the estimated genetic correlation was observed with a mating ratio of one dam per sire, and 10 offspring per sire per environment. When common environmental effects are present in the population and are included in the model, the lowest SE is obtained with mating ratios of at least 5 dams per sire and with a minimum number of 10 offspring per sire per environment. We recommend that studies that aim to estimate the magnitude of G × E in pigs, chicken, and fish should acknowledge the potential presence of common environmental effects and adjust the mating ratio accordingly.

Key words: breeding programs, genetic correlation, genotype by environment interaction, population structure

Introduction

The purpose of animal breeding programs is to select and breed animals that will produce more efficiently under future production settings. How well these animals perform depends on both their genotype and the production environments where they are selected and later kept (Falconer, 1952; Falconer and Mackay, 1996). Ranking of animals may differ between environments as a result of genotype by environment (G × E) interaction (Falconer and Mackay, 1996). The presence of G × E can be determined by estimating the genetic correlation (r_g)

Abbreviations

FS	full-sib
HS	half-sib

between the environments, using measurements of related individuals for a given trait in two environments (Falconer, 1952). Having unbiased and accurate estimates of the genetic correlation is important when predicting the response in another environment than the one where the selection took place and when optimizing genetic improvement programs in terms of collecting phenotypic and genotypic information in production environments (Mulder and Bijma, 2005; Mulder, 2016).

Genetic correlations can be estimated based on records of relatives in two environments. Similarities between related individuals can be not only due to heritable genetic effects but also due to common environmental effects. Estimating the correlation can be done from data collected in breeding programs (Falconer, 1952), from specifically designed experiments, or a combination of both (Sae-Lim et al., 2016). Estimating genetic correlations between environments based on full-sibs (**FSs**) and half-sibs (**HSs**) in different environments is common in pigs, poultry, and fish, while in cattle estimation of genetic correlations between environments is almost entirely based on HSs. FS or HS animals are initially kept together as, for example, litter groups (pigs), hatched chicks in pens (chickens), or hatched fry and fingerlings in tanks or cages (fish) until the age of weaning or individual tagging. This group rearing period potentially leads to common environmental effects. To get an accurate and unbiased estimate of the genetic correlation, this common environmental effect should be taken into account.

Predicting the standard error (**SE**) of the genetic correlation has been studied for many years. Robertson (1959) presented a theoretical basis for predicting the SE of the genetic correlation for specific types of relatives (i.e. either only FS or only HS) in the absence of common environmental effects. Sae-Lim et al. (2010) studied the SE of the estimated genetic correlation for a specific combination of an FS–HS design in which one male was mated to two different females and common environmental effects were not accounted for. Omitting the common environmental effects (c^2) can lead to severely biased estimates of the genetic parameters (Clément et al., 2001), and, therefore, accounting for these effects is of importance when estimating genetic correlations. Bijma and Bastiaansen (2014) presented a formula for the SE of genetic correlation estimates taking common environmental effects into account, but their work was developed for a purebred–crossbred scenario, thereby limiting the family design to only HS groups.

Knowledge on the optimal FS–HS structure to minimize bias and SE of the estimates of the genetic correlation between performance in different environments is currently lacking for situations where common environmental effects are suspected to be present, such as in chicken and fish. The main objective of this study was, therefore, to identify optimal mating designs for the estimation of genetic correlation between environments in the presence of common environmental effects. To achieve this, we used stochastic simulations to compare scenarios where the ratio of FS and HS relationships, as well as the mating ratio, was varied. Results of the simulations were compared with the deterministic equations by Robertson (1959) and Bijma and Bastiaansen (2014).

Materials and Methods

Animal Care and Use Committee approval was not needed because data were simulated.

Experimental populations data were created by stochastic simulations in R software version 3.2.2 (R Development Core Team, 2016) running in RStudio version 1.0.153 (RStudio Team, 2015). The genetic correlation between environments was estimated using an animal model implemented in ASReml (Gilmour et al., 2014b). Simulation was performed with and without the presence of common environmental effects for each FS family, and estimation was performed with and without accounting for common environmental effects, resulting in a 2 × 2 design of presence or absence of common environmental effects.

Populations

The testing structure was based on a split-family design, where the generated FS offspring were divided equally over two environments, had trait records, and only their parents contributed to the one-generation pedigree. To compare scenarios with equal requirements for phenotyping efforts, all our designs had a constant population size of 2,000 individuals per environment. This population size is the same as the starting point for simulations in Sae-Lim et al. (2010), who showed unbiased estimates of r_g for designs with 100 families with 20 offspring. The trait heritability (h^2) was 0.3 in each environment, and the genetic correlation (r_g) between environments was set equal to 0.5. The investigated variables were the number of sires (20 to 1,000), number of dams per sire (1,5, or 10), and the resulting number of offspring per dam was adjusted to keep the total population size constant. All scenarios were simulated with and without common environmental effects and all datasets were analyzed with a model accounting for or ignoring the common environmental effects. The value c^2 is the variance of the common environmental effect (σ^2_C)as a proportion of the total phenotypic variance. The simulated values for c^2 were 0, 0.05, and 0.1, which is in the range of what is commonly observed in livestock and fish (Table 1).

Breeding values and phenotypes

True breeding values of the parents for two environments were drawn from a bivariate normal distribution with means of zero, variance–covariance matrix equal to $\begin{bmatrix} 0.3 & 0.15 \\ 0.15 & 0.3 \end{bmatrix}$, and using a heritability of 0.3 in both environments. The phenotypic variances (σ^2_P) were set to 1 in both environments. The genetic correlation between the two environments was set to 0.5. Common environmental effects in, for example, chicken, pigs, or fish are typically due to common rearing of juveniles before they are exposed to the different growing environments. Therefore, due to these same rearing conditions, the correlation between common environmental effects for the two environments was set to 1. For each FS group, the common environmental effect was equal for all individuals. The common environmental effects for each FS group were simulated by drawing values from a standard normal distribution with a mean of zero and variance equal to the common environmental effect variance ($N(0, \sigma^2_C)$).

A true breeding value for each individual offspring, for each specific environment (i) in which it was kept, was simulated as:

$$TBV_{0i} = 0.5\left(TBV_{mi} + TBV_{fi}\right) + ms_i,$$

Table 1. Simulated population structures for three different mating ratios, each one with three different levels of common environmental effects (c^2)

Mating ratio	c^2	Population structure							
1:2	0; 0.05; 0.1	Number of sires	1,000	500	200	100	50	40	25	20
		HS number per environment[1]	2	4	10	20	40	50	80	100
		Number of dams	2,000	1,000	400	200	100	80	50	40
		FS number per environment	1	2	5	10	20	25	40	50
1:5	0; 0.05; 0.1	Number of sires	400	200	100	80	40	20		
		HS number per environment[1]	5	10	20	25	50	100		
		Number of dams	2,000	1,000	500	400	200	100		
		FS number per environment	1	2	4	5	10	20		
1:10	0; 0.05; 0.1	Number of sires	200	100	50	25	20			
		HS number per environment[1]	10	20	40	80	100			
		Number of dams	2,000	1,000	500	250	200			
		FS number per environment	1	2	4	8	10			

[1]Sire offspring per environment.

where TBV_{mi} and TBV_{fi} are the true breeding values previously assigned to the male and female parents respectively, and ms_i is the Mendelian sampling term. The Mendelian sampling term was drawn from a normal distribution with mean of zero, and a variance equal to half the additive genetic variance ($N(0, 0.5\sigma^2_A)$). Offspring phenotypes were obtained by adding their true breeding value, common environmental effect, and a simulated environmental effect sampled from a normal distribution $N(0, (1 - \sigma^2_A - \sigma^2_C))$. Simulations experiments were replicated 500 times.

Estimation of genetic parameters

The genetic correlation was estimated from the simulated data using two different models, one accounting for the presence of common environmental effects and another ignoring these effects. The models were fitted as follows:

Model 1

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{x}_1 & 0\\ 0 & \mathbf{x}_2\end{bmatrix}\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_1 & 0\\ 0 & \mathbf{Z}_2\end{bmatrix}\begin{bmatrix}\mathbf{a}_1\\ \mathbf{a}_2\end{bmatrix} + \begin{bmatrix}\mathbf{W}_1\\ \mathbf{W}_2\end{bmatrix}\mathbf{c} + \begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix} \quad (1)$$

Model 2

$$\begin{bmatrix}\mathbf{ly}_1\\ \mathbf{ly}_2\end{bmatrix} = \begin{bmatrix}\mathbf{x}_1 & 0\\ 0 & \mathbf{x}_2\end{bmatrix}\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_1 & 0\\ 0 & \mathbf{Z}_2\end{bmatrix}\begin{bmatrix}\mathbf{a}_1\\ \mathbf{a}_2\end{bmatrix} + \begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix} \quad (2)$$

where $\mathbf{y}_1$ and $\mathbf{y}_2$ are vectors with the phenotypes measured in environment 1 and 2, $\mathbf{x}_1$ and $\mathbf{x}_2$ are the incidence vectors relating the traits to the mean in environment 1 (μ_1) or environment 2 (μ_2), $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are the incidence matrices relating the phenotypes to the random additive genetic effect in environments 1 ($\mathbf{a}_1$) and 2 ($\mathbf{a}_2$), $\mathbf{W}_1$ and $\mathbf{W}_2$ are the incidence matrices relating the phenotypes per sire offspring to the common environmental effect ($\mathbf{c}$), and $\mathbf{e}_1$ and $\mathbf{e}_2$ are the vectors containing the random residual effects. The estimated breeding values were assumed to follow a normal distribution ($\sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\mathbf{A}\sigma^2_{A1} & \mathbf{A}\sigma_{A1,2}\\ \mathbf{A}\sigma_{A1,2} & \mathbf{A}\sigma^2_{A2}\end{bmatrix}\right)$), where A is the pedigree relationship matrix, σ^2_{A1} is the genetic variance in environment 1, σ^2_{A2} is the genetic variance in environment 2, and $\sigma_{A1,2}$ is the genetic covariance between environments 1 and 2. The residual covariance was fixed at zero, because each animal is performing in only one environment. The residuals were assumed to be independent ($\sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\mathbf{I}\sigma^2_{e1} & 0\\ 0 & \mathbf{I}\sigma^2_{e2}\end{bmatrix}\right)$), where σ^2_{e1} and σ^2_{e2} are the residual variances in environments 1 and 2, respectively. The residual covariance was fixed at zero, because each animal is performing in only one environment.

Summarizing simulation output

The estimates and SEs were obtained as reported by ASReml version 4.1 software (Gilmour et al., 2014b) for each of the genetic parameters h^2_1, h^2_2, r_g, and c^2. Replicates with estimates of r_g in the –1 to 1 range, that had converged, and where the variance–covariance matrices were positive definite were kept for further analysis. The proportion of replicates that did not converged ranged from 0.2% to 49.2% (Supplementary Table S1). For most scenarios, the SE of the genetic correlation was negatively correlated with the size of the correlation estimate. However, for the same level of correlation, the SEs still differed considerably between replicates; therefore, the SE could not be predicted from the estimate. This illustrates the scenarios cannot be evaluated based on only one replicate (Supplementary Figure S1). The R-script used for the simulations is included as Supplementary Material S1.

Accuracy of genetic correlation estimates

The accuracy of the estimated genetic correlation was obtained by 1) the standard deviation across estimates from different simulation replicates and 2) the average of the approximated SEs reported by ASReml (Fischer et al., 2004; Gilmour et al., 2014a). If the true and statistical models are the same, it is expected that the standard deviation across estimates is similar to the average SE if the approximation is accurate. For scenarios without a common environmental effect, the SEs of the genetic correlation were compared with deterministic predictions, which assume that the true parameters are known. First, for scenarios with only FSs, or only HSs, the SE was predicted by Robertson's equation (1959):

$$SE\left(\hat{r}_g\right) \approx \sqrt{\frac{\left[1 + nt\left(1 - r^2_g\right)\right]^2 + r^2_g}{(N-1)\,n^2t^2}} \quad (3)$$

Where N is the number of sire families, n is the number of offspring per environment, and t is the intra-class correlation (e.g., for HSs, $t = 0.25h^2$), and r_g is the known genetic correlation between environments. Second, for scenarios with only HSs, the SE was predicted using the equation from Bijma and Bastiaansen (2014):

$$SE(\hat{r}_g) \approx \sqrt{\frac{\frac{1}{\rho_{x^2}\rho_{y^2}} + \left(1 + \frac{0.5}{\rho_{x^4}} + \frac{0.5}{\rho_{y^4}} - \frac{2}{\rho_{x^2}} - \frac{2}{\rho_{y^2}}\right)r^2_g + r^4_g}{N-1}} \quad (4)$$

Where N is the number of sire families, ρ_x^2 and ρ_y^2 are the reliabilities of estimated breeding values in each environment calculated using formulas 2 and 3 in Bijma and Bastiaansen (2014), and r_g is the true genetic correlation between environments.

Results

The average SEs as reported by ASReml were not different from the standard deviation estimated over replicates. Therefore, we only report the latter. The SE of the $\hat{r}_g$ for the scenarios where equations from Robertson (1959) and Bijma and Bastiaansen (2014) could be applied (i.e., only FS or only HS) did not differ from SE obtained by the simulations (see Supplementary Table S2 for the SEs). Only results from simulations are reported here.

SE of estimated r_g with $c^2 = 0$

In the absence of common environmental effects, the smallest SE was obtained with FS groups only, where the minimum SE for our population size of 2,000 per environment (0.097) was found for 10 offspring per sire per environment (Figure 1). With HS groups only, the minimum SE (0.144) was obtained with 20 offspring per sire per environment (Figure 1).

In the absence of common environmental effects, having an FS–HS structure gave higher SE of the $\hat{r}_g$ than having only FS, but lower SE than having only HS (Figures 1 and 2a). With 10 offspring per sire per environment, the SE of the $\hat{r}_g$ increased from 0.097 in a 1:1 mating ratio (Figure 1) to 0.100, 0.114, and 0.128 for a 1:2, 1:5, and 1:10 ratio, respectively (Figure 2a). When the number of offspring per sire per environment increased, and consequently the number of sires decreased, the SE increases rapidly with a 1:1 mating ratio (Figure 1) but much slower with more than one dam per sire (Figure 2a). When the number of offspring per sire per environment was equal or greater than 40, the SE for a 1:1 mating ratio (FS family design) was larger than the SE for all the other mating ratios (Figure 2a). In summary, in the absence of common environmental effects, a 1:1 mating ratio with no more than 10 offspring per sire per environment will result in the smallest SE.

SE of estimated r_g with $c^2 > 0$

In the presence of c^2 in the population, and when accounting for c^2 in the model, the smallest SE of the $\hat{r}_g$ for all mating ratios was found when the number of offspring per sire per environment was equal to 20 (Figure 2b and c). The lowest SE of the $\hat{r}_g$ was reached when the mating ratio was 1:10 with

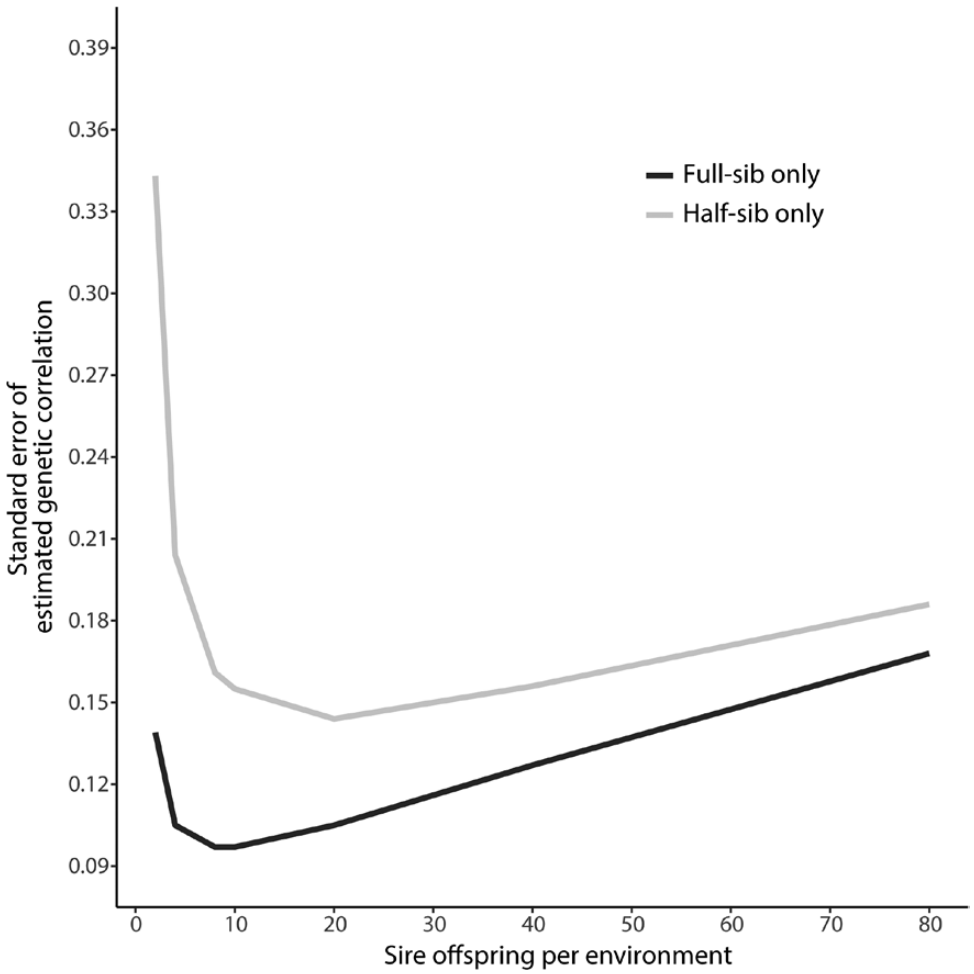

Figure 1. SE of the estimated genetic correlation for simulated scenarios for only HS and only FS for a population size of 2,000 individuals per environment and no common environmental effects (c^2).

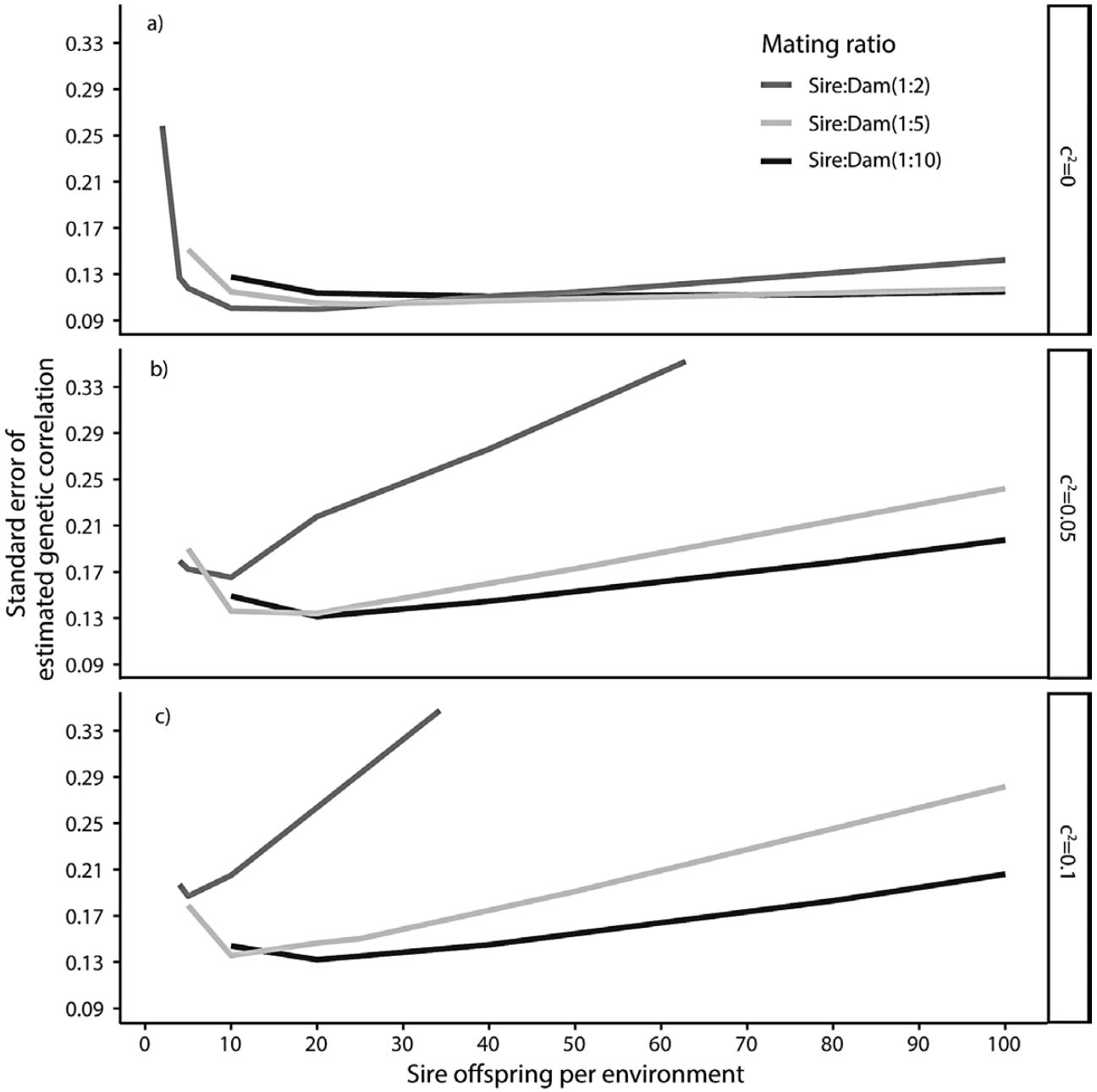

Figure 2. SE of the estimated genetic correlation for different simulated scenarios using the correct model for three mating ratios for a population size of 2,000 individuals per environment, and three levels of common environmental effects (c^2). In panel (a), $c^2 = 0$ and no common environmental effects included in the model, in panel (b) $c^2 = 0.05$ and the common environmental effects are included in the model, and in panel (c) $c^2 = 0.1$ and the common environmental effects are included in the model.

20 offspring per sire per environment, which resulted in each offspring having 1 FS and 18 HS in the same environment (SE = 0.131; Figure 2b and c; Table 1). When c^2 was present, and included in the model, the SEs of the $\hat{r}_g$ increased for all scenarios, compared with the scenarios when c^2 was not present and not included in the model (Figure 2). When the number of offspring per sire per environment increased, the SE for the 1:2 mating ratio became extremely large. However, when there were more females mated per male, the SE of the $\hat{r}_g$ also increased when the number of offspring per sire per environment increased, but at a much lower rate (Figure 2). In the presence of c^2, the best design with smallest SE is of 10 females mated per male with 20 offspring per sire per environment.

Estimates of r_g and c^2 from the correct model

In the absence of c^2 and without the common environmental effect in the model, the $\hat{r}_g$ were unbiased (Figure 3c). When $c^2 > 0$, and the common environmental effect was included in the model, the $\hat{r}_g$ were also unbiased (for all different mating ratios and levels of c^2 tested (Figure 3b; only showing results for 1:10 mating ratio). The estimates of c^2 themselves were also unbiased when common environmental effects were present in the population and included in the model (Figure 4b).

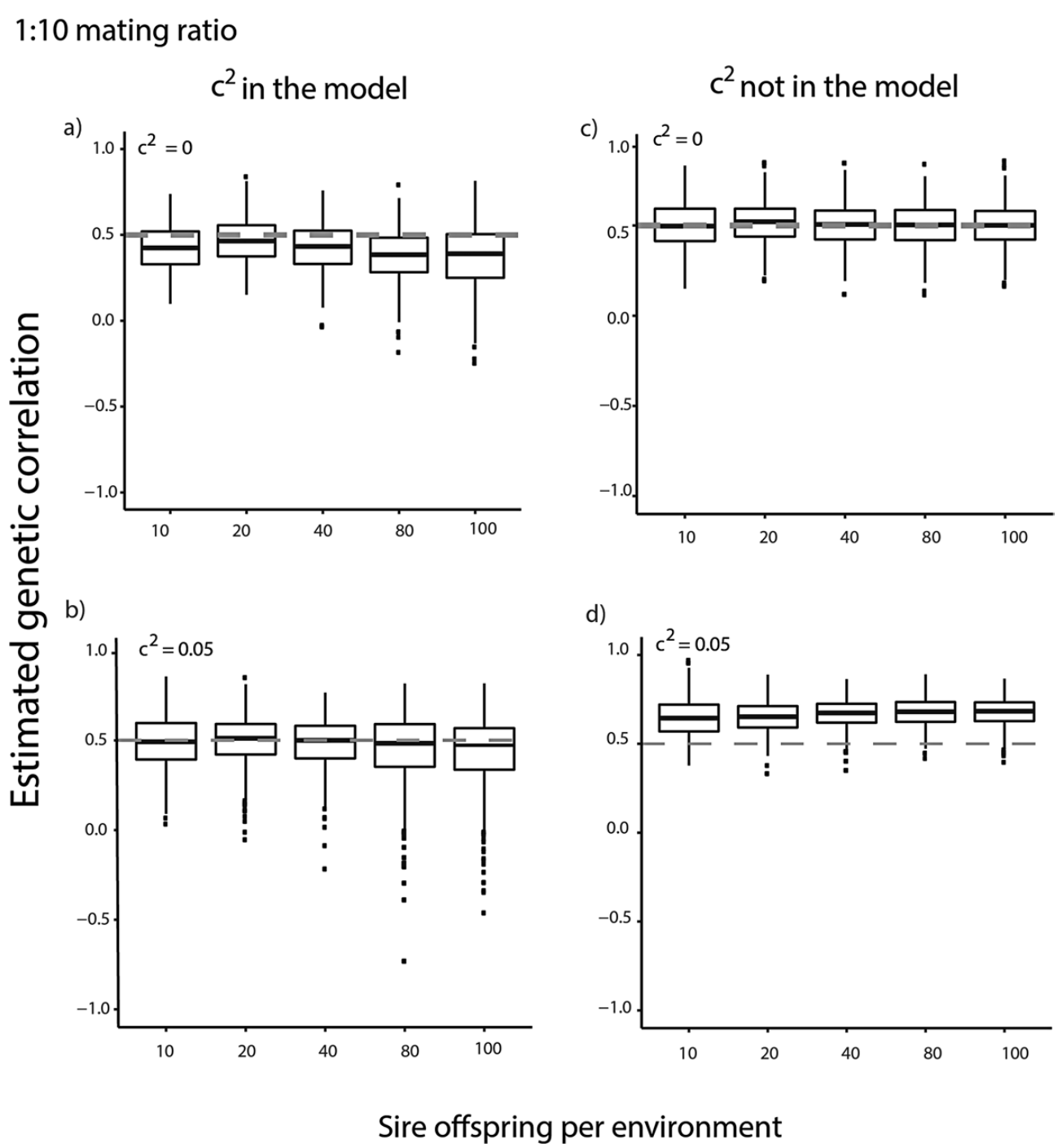

Figure 3. Distribution of the estimated genetic correlation for 1:10 mating ratio for a population size of 2,000 individuals per environment when (a) $c^2 = 0$ and (b) $c^2 = 0.05$ and common environmental effects are included in the model. When (c) $c^2 = 0$ and (d) $c^2 = 0.05$ and common environmental effects are not included in the model. The dashed line indicates the simulated value of the genetic correlation (r_g=0.5).

Estimates of r_g and c^2 from the incorrect model

In the absence of common environmental effects in the population ($c^2 = 0$) but with the common environmental effects present in the model, there was a downward bias of the $\hat{r}_g$ of on average 0.1 (Figure 3a), and an increase in the standard deviation of the genetic correlation estimates. The estimates of c^2 were biased upward by on average 0.02 (Figure 4a). On the other hand, estimates of genetic correlations showed an upward bias of on average 0.3 (from 0.5 to 0.8) when common environmental effects were present in the population, but not included in the model (Figure 3d). We only show the results for the 1:10 mating ratio but this pattern was the same for all different mating ratios and levels of c^2 tested.

Discussion

The aim of this study was to identify the optimal mating design to estimate the genetic correlation between environments in

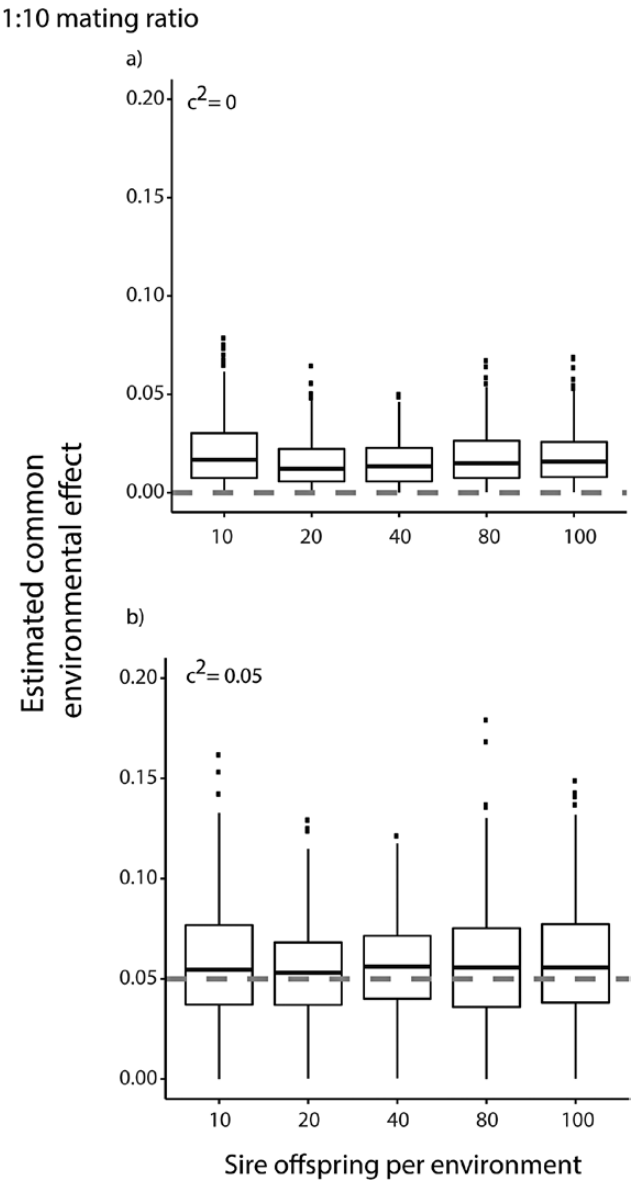

Figure 4. Distribution of the estimated common environmental effect for 1:10 mating ratio for a population size of 2,000 individuals per environment when (a) $c^2 = 0$ and when (b) $c^2 = 0.05$. The dashed line indicates the simulated value of the common environmental effect $c^2 = 0.0$, and $c^2 = 0.05$, respectively.

the presence of common environmental effects and different FS and HS ratios. Estimates of genetic correlations of traits between different environments are of crucial importance for the design and optimization of breeding programs. Neglecting the existence of a G × E interaction will lead to reduced response to selection (Mulder and Bijma, 2005; Dominik and Kinghorn, 2008). For accurate estimates, experimental designs should be optimized. In many situations, the populations used to collect data can be designed to a certain degree.

Some studies can or have to use field data to estimate genetic correlations (Mulder et al., 2004; Haile-Mariam et al., 2015; Sevillano et al., 2016; Godinho et al., 2018), but other studies use experimental designs under specific population structures to estimate the magnitude of G × E (Lwelamira, 2012; Trọng et al., 2013; Sae-Lim et al., 2014; Omasaki et al., 2016; Dottavio et al., 2019; Lillehammer et al., 2019). As common environmental effects generate similarities between individuals, these effects should be taken into account when designing the best population structure to estimate the genetic correlations (Winkelman and Peterson, 1994). Here we show the importance of choosing the experimental design when taking common environmental effects into account. Not accounting for these effects will lead to biased estimates of the genetic correlation and common environmental effects will increase the SE of genetic correlations.

Levels of c^2

Siblings often share a common environment during the early rearing stage, increasing the phenotypic covariance between relatives. In livestock, some studies have reported high levels of c^2 ranging from 0.06 to 0.5. For example, in several pig breeds, the proportion of variance due to common environmental effects has been reported to be around 0.34 for intramuscular fat, between 0.06 and 0.53 for fat composition in Iberian pig lines (Ibáñez-Escriche et al., 2016) and between 0.22 and 0.26 for days to 100 kg (Li and Kennedy, 1994). For catfish body weight,

estimates reported range between 0.06 and 0.33 (Tran, 2016). For Venda chicken body weight, common environmental effects for weight at hatching and weight at 4 wk were reported to be 0.39 and 0.18, respectively (Norris and Ngambi, 2006). Here we show that the higher the level of c^2, the greater the bias of the estimated genetic correlation when not including common environmental effects in the model. Therefore, the high levels of c^2 in these studies highlight the relevance to take common environmental effects into account in the model and to optimize the design for estimating genetic correlations.

Optimizing mating designs

When the number of animals to be phenotyped is fixed, the SE of the estimated genetic correlation can still be influenced by the mating design. In animal breeding programs, nested designs are traditionally used, where one sire is mated to a different number of dams. Bijma and Bastiaansen (2014) predicted the SE of the estimated genetic correlation for mating ratios that are typically used for four species under different levels of c^2. Comparing mating ratios of 1:2 for tilapia, 1:7 for laying hens, 1:10 for pigs, and 1:12 for broilers, they report the highest SE of the estimated genetic correlation for a 1:2 design such as used in Nile tilapia. Increasing the number of dams per sire was found to lower the SE of the $\hat{r}_g$. This is in agreement with our study, where we show that having a higher number of dams per sire will give more accurate estimates of the genetic correlation regardless of the level of c^2. Even though having more than 10 dams per sire may give even smaller SE of the $\hat{r}_g$, this will reduce the number of sires used considerably, which is not desired for reasons of genetic diversity, when the experiment is part of a breeding program.

With a given population size of, for example, 2,000 individuals per environment, having a higher number of females per male reduces the number of FSs (offspring per dam) per environment but keeps the number of HSs (offspring per sire) per environment the same. Distinguishing common environmental effects from genetic effects requires that related individuals exist in different environments, and that there is variation in the degree of relatedness between individuals that share the particular common environment (Kruuk and Hadfield, 2007). To minimize bias in the estimates of the genetic correlation, having more dams per sire should be implemented.

In cases when the number of offspring per sire per environment increases, the number of parents linking this offspring to each environment decreases. Therefore, when the number of offspring per sire is larger than 40, the number of dams per sire should be increased to be at least 5. With common environmental effects present in the population, the optimum number of offspring per sire per environment is equal to 20. Having more dams per sire, at least 5, will then be needed to compensate for the low number of sires.

In accordance with Robertson (1959), in the absence of common environmental effects, the 1:1 mating ratio gives the lowest SEs for the estimated genetic correlation. However, to benefit from using a 1:1 mating ratio, there has to be certainty about the absence of common environmental effects, so that there is no bias on the estimated genetic correlation. Under this mating ratio, the variance between FSs is half of the additive genetic variance plus the variance due to common environmental effects. Ignoring the common environmental effects will, therefore, overestimate the additive genetic variance by two times the common environmental effect variance (Kruuk and Hadfield, 2007). On the other hand, using only HSs gives the highest SEs for the estimated genetic correlation. Formulas and simulations assume a balanced design, and that the fixed effects are known, while the actual estimation of the genetic correlation typically involves unbalanced data and estimation of the fixed effects. It should be noted that the SEs of the estimated genetic correlation obtained from the Bijma and Bastiaansen (2014) formula are slightly lower for all the different population structures compared with the SEs obtained by the simulations and by the Robertson (1959) formula. Therefore, the formula presented by Bijma and Bastiaansen (2014) can be interpreted as a lower bound of the SE of the $\hat{r}_g$ for only HS relationships.

Nested designs, with a 1:2 mating ratio, are commonly seen in, for instance, Nile tilapia breeding programs. These designs are applied because FSs need to be reared separately until tagging before they are transferred to communal growth environment. Family production in some fish breeding programs like tilapia relies on natural mating and families are kept separate for a long period, making it likely that common environmental effects are introduced (Winkelman and Peterson, 1994; Trọng et al., 2013). Two dams per sire is the minimum design to allow the estimation of c^2. However, some G × E studies in Nile tilapia have shown difficulties when trying to estimate genetic parameters while including common environmental effects in the model. They attribute this problem to the low number of dams mated to each sire (1:2 or 1:3 mating ratio) and to a fixed effect such as spawning date being confounded with the c^2 (Omasaki et al., 2016). Other studies have estimated significant levels of common environmental effects present for different tilapia species (Thodesen et al., 2013; Trọng et al., 2013; Thoa et al., 2015). Using a 1:2 mating design in Red Tilapia, estimates for c^2 ranging from 0.23 to 0.59 were found in fish kept in ponds, and estimates between 0.1 and 0.31 were found for fish kept in cages (Nguyen et al., 2017). These significant levels of c^2 and the fact that different studies had trouble in estimating c^2 highlight the importance of optimizing designs to estimate both c^2 and r_g.

With an appropriate experimental design to estimate the presence of common environmental effects, the question arises of when to include or not the common environmental effects in the model. With small estimates of common environmental effects, in the range of 0.01 to 0.02, a logical conclusion could be to remove it from the model. The downside of considering these range of effects being close to 0, or very small, and removing them from the model, can still lead to a bias in the estimates of the genetic correlation and heritabilities. The population structure will also play a role in this decision. Based on the scenario presented here with a constant population size of 2,000 offspring per environment, the common environmental effects should be always included in the model when these common environmental effects are likely to be present based on the husbandry system and biology of the trait.

Impact of biased r_g estimates

Based on the scenario presented here of a constant population size of 2,000 individuals per environment, and a true genetic correlation equal to 0.5, if common environmental effects are present in the population, and are not considered in the model will lead to an overestimation of the genetic correlation of on average 0.3. Biased genetic parameter estimates will lead to an erroneous conclusion about the genetic values of individuals for the different environments, to over- or underestimating the importance of G × E, and possibly suboptimally designed breeding programs. Biased genetic correlation estimates can also lead to erroneous conclusions about the genetic improvement in breeding programs (Dominik and Kinghorn, 2008; Chu et al., 2018). Overestimated genetic correlations

would result from wrongly ignoring the presence of c^2, and could lead to the conclusion that G × E is absent, when in fact it exists. These biased estimates of genetic correlations may cause the decision to have a single breeding program to serve the different environments, while the optimal decision would be to have multiple separate breeding programs (Mulder et al., 2006).

In cases where the genetic correlation is estimated between traits measured in the same individuals, ignoring the common environmental effects is likely to lead to bias similar to what is observed for the G × E scenario presented in this study. In those estimates, the common environmental effects are more likely to have a non-unit correlation, as some traits are more affected by maternal care or other early-life conditions than others. As presented in this study, with a correlation of 1 between common environmental effects, under a 2,000 population size scenario, the estimates of the genetic correlation are severely biased. The extent of the bias may vary with the level of common environmental effects, and with the correlation between effects in different environments, but the bias is not expected to disappear under certain conditions except when the common environmental effects are truly zero.

Robertson (1959) presented a formula to predict the SE of the genetic correlation for restricted family groups, either only FSs or only HSs in the absence of common environmental effects. Later on, Bijma and Bastiaansen (2014) presented a formula to estimate the SE of the genetic correlation accounting for common environmental effects but restricted the family design to only-HS groups. In animal breeding, it is common to estimate the genetic correlation between environments for a combination of FS and HS separated in two environments. With these family designs, the presence of common environmental effects is very likely. Therefore, the previous theoretical bases should be considered with some caveats, as they do not resemble the actual genetic improvement programs. This study acknowledges the combination of FS and HS groups and the presence of common environmental effects for the estimation of the SE of the genetic correlation. This will allow for animal breeders to decide on the appropriate population structure needed to achieve unbiased estimates under more realistic scenarios.

Currently, genomic information is widely implemented to estimate genetic parameters. A study that compared the use of pedigree vs. genomic relationships to estimate the genetic correlation between different traits showed that there was an increase in the accuracy of the estimated genetic correlation when using only genomic relationships. Combining both pedigree and genomic relationships gave the most accurate estimates of the genetic variance and genetic correlations. They conclude that the combination of both relationships can be beneficial for small data sets (Veerkamp et al., 2011). Optimal designs that use genomic relationships to estimate the genetic correlation in the presence of common environmental effects can show different results than the ones illustrated here. We encourage this to be investigated.

This simulation study focused on finding the optimal population design to minimize the SE of the genetic correlation in the presence of common environmental effects. The constant population size of 2,000 individuals per environment allowed us to make comparisons across different mating ratios. We show that mating structures with more than two dams per sire should be used to better disentangle the common environmental effects from the genetic effects, thereby obtaining a lower SE of the estimated genetic correlation. Only when there is certainty about the absence of common environmental effects in the population, a single dam per sire will give the lowest SE and an unbiased estimate of the genetic correlation.

Supplementary Data

Supplementary data are available at *Journal of Animal Science* online.

Acknowledgment

The authors sincerely thank the Koepon Foundation for providing a scholarship to the first author.

Conflict of interest statement

The authors declare no conflict of interest.

Literature Cited

Bijma, P., and J. W. Bastiaansen. 2014. Standard error of the genetic correlation: how much data do we need to estimate a purebred-crossbred genetic correlation? *Genet. Sel. Evol.* **46**:79. doi:10.1186/s12711-014-0079-z

Chu, T. T., J. W. M. Bastiaansen, E. Norberg, and P. Berg. 2018. On farm observations to increase genetic gain in breeding schemes for village poultry production – a simulation study. *Acta Agri. Scand. AAnim. Sci.* **68**:1–10. doi:10.1080/09064702.09062018.01543444

Clément, V., B. Bibé, É. Verrier, J.-M. Elsen, E. Manfredi, J. Bouix, and É. Hanocq. 2001. Simulation analysis to test the influence of model adequacy and data structure on the estimation of genetic parameters for traits with direct and maternal effects. *Gen. Sel. Evol.* **33**:369. doi:10.1186/1297-9686-1133-1184-1369

Dominik, S., and B. Kinghorn. 2008. Neglecting genotype × environment interaction results in biased predictions from selection index calculations. *Livest. Sci.* **114**:233–240. doi:10.1016/j.livsci.2007.1005.1004

Dottavio, A., S. Advínculo, A. Martines, J. Librera, Z. Canet, B. Romera, and R. Di Masso. 2019. Interacción genotipo × estación del año sobre caracteres de producción de carne en pollos camperos. *Compendio de Ciencias Veterinarias*. **9**:15–21. doi:10.18004/compend.cienc.vet.2019.09.01.15-21

Falconer, D. S. 1952. The problem of environment and selection. *Am. Nat.* **86**:293–298. doi:10.1086/281736

Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to quantitative genetics*. London: Longman Group.

Fischer, T. M., A. R. Gilmour, and J. H. van der Werf. 2004. Computing approximate standard errors for genetic parameters derived from random regression models fitted by average information REML. *Genet. Sel. Evol.* **36**:363–369. doi:10.1051/gse:2004006

Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2014a. *ASReml user guide Release 4.2*. Hemel Hempstead, UK: VSN International Ltd. www.vsni.co.uk

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2014b. *ASReml user guide Release 4.1 structural specification*. Hemel Hempstead, UK: VSN International Ltd.

Godinho, R., R. Bergsma, F. Silva, C. Sevillano, E. Knol, M. Lopes, P. Lopes, J. Bastiaansen, and S. Guimarães. 2018. Genetic correlations between feed efficiency traits, and growth performance and carcass traits in purebred and crossbred pigs. *J. Anim. Sci.* **96**:817–829. doi:10.1093/jas/skx1011

Haile-Mariam, M., J. Pryce, C. Schrooten, and B. Hayes. 2015. Including overseas performance information in genomic evaluations of Australian dairy cattle. *J. Dairy Sci.* **98**:3443–3459. doi:10.3168/jds.2014–8785

Ibáñez-Escriche, N., E. Magallón, E. Gonzalez, J. F. Tejeda, and J. L. Noguera. 2016. Genetic parameters and crossbreeding

effects of fat deposition and fatty acid profiles in Iberian pig lines. *J. Anim. Sci.* **94**:28–37. doi:10.2527/jas.2015–9433

Kruuk, L. E., and J. D. Hadfield. 2007. How to separate genetic and environmental causes of similarity between relatives. *J. Evol. Biol.* **20**:1890–1903. doi:10.1111/j.1420-9101.2007.01377.x

Li, X., and B. W. Kennedy. 1994. Genetic parameters for growth rate and backfat in Canadian Yorkshire, Landrace, Duroc, and Hampshire pigs. *J. Anim. Sci.* **72**:1450–1454. doi:10.2527/1994.7261450x

Lillehammer, M., S. A. Boison, A. Norris, M. Løvoll, H. Bakke, and B. Gjerde. 2019. Genetic parameters of resistance to amoebic gill disease in two Norwegian Atlantic salmon populations. *Aquaculture* **508**:83–89. doi:10.1016/j.aquaculture.2019.1004.1072

Lwelamira, J. 2012. Genotype-environment (GXE) interaction for body weights for Kuchi chicken ecotype of Tanzania reared under intensive and extensive management. *Global Journal of Medical Research*. **12**:57–65. doi:10.3923/ijps.2012.96.102

Mulder, H. A. 2016. Genomic selection improves response to selection in resilience by exploiting genotype by environment interactions. *Front. Genet.* **7**:178. doi:10.3389/fgene.2016.00178

Mulder, H. A., and P. Bijma. 2005. Effects of genotype × environment interaction on genetic gain in breeding programs. *J. Anim. Sci.* **83**:49–61. doi:10.2527/2005.83149x

Mulder, H. A., A. F. Groen, G. De Jong, and P. Bijma. 2004. Genotype × environment interaction for yield and somatic cell score with automatic and conventional milking systems. *J. Dairy Sci.* **87**:1487–1495. doi:10.3168/jds.S0022-0302(04)73300-7

Mulder, H. A., R. F. Veerkamp, B. J. Ducro, J. A. van Arendonk, and P. Bijma. 2006. Optimization of dairy cattle breeding programs for different environments with genotype by environment interaction. *J. Dairy Sci.* **89**:1740–1752. doi:10.3168/jds.S0022-0302(06)72242-1

Nguyen, N. H., A. Hamzah, and N. P. Thoa. 2017. Effects of genotype by environment interaction on genetic gain and genetic parameter estimates in Red Tilapia (Oreochromis spp.). *Front. Genet.* **8**:82. doi:10.3389/fgene.2017.00082

Norris, D., and J. W. Ngambi. 2006. Genetic parameter estimates for body weight in local Venda chickens. *Trop. Anim. Health Prod.* **38**:605–609. doi:10.1007/s11250-006-4420-6

Omasaki, S. K., H. Charo-Karisa, A. K. Kahi, and H. Komen. 2016. Genotype by environment interaction for harvest weight, growth rate and shape between monosex and mixed sex Nile tilapia (*Oreochromis niloticus*). *Aquaculture* **458**:75–81. doi:10.1016/j.aquaculture.2016.1002.1033

R Development Core Team. 2016. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. [accessed December 8, 2016]. http://www.R-project.org.

Robertson, A. 1959. The sampling variance of the genetic correlation coefficient. *Biometrics* **15**:469–485. doi:10.2307/2527750

RStudio Team. 2015. RStudio: integrated development for R. Boston (MA): RStudio, Inc. [accessed June 30, 2015]. http://www.rstudio.com/.

Sae-Lim, P., B. Gjerde, H. M. Nielsen, H. Mulder, and A. Kause. 2016. A review of genotype-by-environment interaction and micro-environmental sensitivity in aquaculture species. *Rev. Aquacult.* **8**:369–393. doi:10.1111/raq.12098

Sae-Lim, P., H. Komen, and A. Kause. 2010. Bias and precision of estimates of genotype-by-environment interaction: a simulation study. *Aquaculture* **310**:66–73. doi:10.1016/j.aquaculture.2010.1010.1020

Sae-Lim, P., H. Komen, A. Kause, and H. A. Mulder. 2014. Identifying environmental variables explaining genotype-by-environment interaction for body weight of rainbow trout (*Onchorynchus mykiss*): reaction norm and factor analytic models. *Genet. Sel. Evol.* **46**:16. doi:10.1186/1297-9686-46-16

Sevillano, C. A., H. A. Mulder, H. Rashidi, P. K. Mathur, and E. F. Knol. 2016. Genetic variation for farrowing rate in pigs in response to change in photoperiod and ambient temperature. *J. Anim. Sci.* **94**:3185–3197. doi:10.2527/jas.2015-9915

Thoa, N. P., W. Knibb, N. H. Ninh, N. Van Dai, P. H. Nhat, L. M. Toan, and N. H. Nguyen. 2015. Genetic variation in survival of tilapia (*Oreochromis niloticus*, Linnaeus, 1758) fry during the early phase of rearing in brackish water environment (5–10ppt). *Aquaculture* **442**:112–118. doi:10.1016/j.aquaculture.2015.1002.1040

Thodesen, J., M. Rye, Y.-X. Wang, S.-J. Li, H. B. Bentsen, and T. Gjedrem. 2013. Genetic improvement of tilapias in China: genetic parameters and selection responses in growth, pond survival and cold-water tolerance of blue tilapia (*Oreochromis aureus*) after four generations of multi-trait selection. *Aquaculture* **396–399**:32–42. doi:10.1016/j.aquaculture.2013.1002.1010

Tran, P. H. 2016. *Analysis of crosses between three strains of striped catfish (Pangasianodon hypophthalmus) in Viet Nam* [master's thesis]. Ås, Norway: Norwegian University of Life Sciences.

Trọng, T. Q., H. A. Mulder, J. A. M. van Arendonk, and H. Komen. 2013. Heritability and genotype by environment interaction estimates for harvest weight, growth rate, and shape of Nile tilapia (*Oreochromis niloticus*) grown in river cage and VAC in Vietnam. *Aquaculture* **384–387**:119–127. doi:10.1016/j.aquaculture.2012.1012.1022

Veerkamp, R. F., H. A. Mulder, R. Thompson, and M. P. Calus. 2011. Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. *J. Dairy Sci.* **94**:4189–4197. doi:10.3168/jds.2011-4223

Winkelman, A. M., and R. G. Peterson. 1994. Heritabilities, dominance variation, common environmental effects and genotype by environment interactions for weight and length in chinook salmon. *Aquaculture* **125**:17–30. doi:10.1016/0044-8486(1094)90278-X