Genetic correlation: a heritable parameter

B.C.D. Cuyabano[1]

[1] *Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France*

Corresponding author: beatriz.castro-dias-cuyabano@inrae.fr

Abstract

Breeding programs rely on selection of individuals through their breeding values to simultaneously improve multiple traits of commercial value. In order to adequately select candidates to breed for a next generation, the genetic relationships between traits are considered in the selection index that summarizes all the traits for each selection candidate. The methods deployed in genetic evaluations rely strongly in gaussian distributions describing the data, and consider the genetic relationships between traits in the form of genetic correlations determining the joint distribution of breeding values from different traits. In this manner, genetic correlations are treated as parameters, estimated on a base population for reference. However, genetic correlations depend on the involved traits' architecture, thus depending on the genotype presented by each individual, and therefore, different individuals may present different potential for genetic correlations. Moreover, different potential for genetic correlations may partially represent a latent physiological trait responsible to balance the phenotypic expression of the measurable production traits. In practice, individual-specific genetic correlations (iSGC) can be obtained for individuals with many phenotyped descendants, as the expressed genetic correlation between the estimated breeding values among their offspring. Since the expressed iSGC depends on the involved traits' genetic architecture, part of an individual's iSGC can be transmitted to the offspring. In order to study the heritability of iSGC, two-trait genetic evaluations were performed on every pairwise combination of five traits from a French Holstein dairy cattle population: milk and protein yield (MY and PY), milking speed (MSPD), somatic cell score (SCS), and conception rate (CR). The iSGC between every pair of the five traits were obtained for ~1200 bulls with more than 500 phenotyped daughters in this population, and these iSGC were each evaluated as a phenotype with a single-trait model. This study confirmed the hypothesis that genetic correlations, when expressed as iSGC, are heritable parameters, with significant heritabilities ranging from 0.11 (iSGC between SCS and CR) to 0.51 (iSGC between PY and SCS).

Key words: Selection index, Multi-trait genetic evaluation, Genetic trade-off, Dairy cattle, Latent phenotype, Physiological traits

Introduction

Breeding programs aim to select for multiple commercial traits, in order to achieve genetic progress for all of them. Many of these traits are genetically correlated, and a negative correlation means that an antagonism between two traits exists. In dairy cattle, the genetic trade-off often lies between production and either fertility or health traits (Boichard & Manfredi, 1994; Pryce et al., 1997; Rauw et al., 1998; Roxström et al., 2001; Windig et al., 2006). Therefore, in order to avoid that selection for one trait is detrimental to the other (Hazel et al., 1994), selection must account for these negative correlations. This is typically done through a selection index, *i.e.*, a linear combination of traits, whose weights are defined by, among other information, the genetic correlations between the traits involved (Hazel, 1943; Hazel et al., 1994; Miglior et al., 2017).

Genetic correlations between traits are considered a populational parameter that

defines the joint normal distribution imposed to the breeding values in genetic evaluations. However, in this manner, genetic correlations are assumed equal to all individuals, an assumption that ignores the fact that different individuals may present different physiological trade-off regulation between traits (Berry et al., 2016; Cuyabano et al., 2024). This hypothesis has been revisited by Cuyabano et al. (2024), who, in a study of the trade-off between production and fertility in the French Montbéliarde population, have shown that different sires could express different genetic correlations through their daughters, between these traits.

Because the study of Cuyabano et al. (2024) had only 247 sires with enough daughters evaluated so that reliable genetic correlations could be obtained at the individual level (*i.e.* for each sire), no further inferences could be drawn, with respect to the genetic background of these *individual-specific genetic correlations* (iSGC).

This current study hypothesized that if the different genetic correlations expressed by different sires are simply a feature of recombination and different allele frequencies in different family lines, then none or very weak heritabilities are expected to be observed for the iSGC. However, if the iSGC represent, even if only partially, a latent physiological phenotype, non-zero heritabilities should be observed for the iSGC.

To support this hypothesis that non-zero heritabilities associated to the iSGC may suggest their representation of a latent physiological phenotype, simulations were deployed. Breeding values were simulated for multiple traits, with their genetic correlations solely due to pleiotropic QTL and linkage disequilibrium between non-pleiotropic sites, in order to show that when no physiological trait was involved in the differences between genetic correlations, no heritability was captured by the iSGC.

For the real data analysis, this current study up-scaled the work from Cuyabano et al. (2024), by calculating iSGC for 1161 sires from a French Holstein dairy cattle population, between each pair of five traits of commercial interest (milk and protein yield, milking speed, cow conception rate, and somatic cell score). Heritabilities were then estimated for the iSGC, under the hypothesis that non-zero estimates suggest the representation of a latent phenotyped through the iSGC.

Materials and Methods

Bi-variate genetic evaluation model

Two-trait animal models were deployed for the genetic evaluations in this study, given by:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}, \qquad (1)$$

in which y_1 and y_2 are the vectors of phenotypes for traits 1 and 2 respectively; $g_1 \sim N(0, A\sigma^2_{g_1})$ and $g_2 \sim N(0, A\sigma^2_{g_2})$ are the vectors of breeding values for these two traits, with $Cov(g_1, g_2)$=$A\sigma_{g_{12}}$, such that A is the pedigree relationship matrix; $\sigma^2_{g_1}$ and $\sigma^2_{g_2}$ are the additive genetic variances, and $\sigma_{g_{12}}$ is the genetic covariance between the two traits; $\varepsilon_1 \sim N(0, I_n\sigma^2_{\varepsilon_1})$ and $\varepsilon_2 \sim N(0, I_n\sigma^2_{\varepsilon_2})$ are the random residuals, with $Cov(\varepsilon_1, \varepsilon_2)$=$I_n\sigma_{\varepsilon_{12}}$; $\sigma^2_{\varepsilon_1}$ and $\sigma^2_{\varepsilon_2}$ are the residual variances, and $\sigma_{\varepsilon_{12}}$ is the residual covariance.

The genetic evaluation model in equation (1) was implemented under a Bayesian framework, using the GIBBS3F90 module from the BLUPF90 family of (Misztal et al., 2018), with the software's default prior distributions for the breeding values and (co)variance parameters. A total of 300,000 samples were generated, with the first 100,000 discarded as burn-in. On the remaining 200,000 samples, a thinning parameter of 200 iterations was applied, resulting in 1000 effective samples used to compute the estimated breeding values (EBV) and (co)variance parameters. To assist convergence of the (co)variance parameters, initial values were provided, using the current genetic (co)variances used for these five traits

in the French national genetic evaluation. Convergence was assessed visually through plots of the 1000 effective samples for the genetic (co)variances.

Heritabilities (h_1^2 and h_2^2) and genetic correlations (ρ_{12}) between the traits were obtained from the estimated (co)variance parameters, as:

$$\hat{h}_1^2 = \frac{\hat{\sigma}_{g_1}^2}{\hat{\sigma}_{g_1}^2 + \hat{\sigma}_{\varepsilon_1}^2} \text{ and } \hat{h}_2^2 = \frac{\hat{\sigma}_{g_2}^2}{\hat{\sigma}_{g_2}^2 + \hat{\sigma}_{\varepsilon_2}^2}, \quad (2)$$

$$\hat{\rho}_{12} = \frac{\hat{\sigma}_{g_{12}}}{\hat{\sigma}_{g_1}\hat{\sigma}_{g_2}}. \quad (3)$$

Real data

The dairy cattle data used for the present study was from the French Holstein population. The bi-variate genetic evaluations were implemented for every pair of the following five traits: milk and protein yield (MY and PY), milking speed (MSPD), somatic cell score (SCS), and cow conception rate (CR), measured as artificial insemination's success/failure on lactating cows (*i.e.* heifers excluded). The phenotypes entered for the evaluations performed in this study were in the form of yield deviations (YD), issued from the French national genetic evaluation, which evaluates MY, PY, SCS as 305-day phenotypes corrected for the duration; performance records comprise all lactations records per cow, and the model accounts for the repeatability (*i.e.*, for the permanent environment of the cow). A total of 4,501,624 cows born between 1991-2020 had YD deviations available for all five traits, with a pedigree file containing a total of 8,275,018 animals that traced back three generations from the cows with performances.

Simulated data

The simulated data consisted of ten replicates of populations with a founder population followed by 30 generations under selection. Generations were non-overlapping, each with 1000 individuals, among them 200 males and 800 females. Selection was performed at each generation for the top 20% males, based on a selection index build from their true simulated breeding values, assuming equal weights for all simulated traits. Pedigree information was kept for the simulated populations.

Five traits were simulated with additive effects associated to them, and genetic correlations were solely due to pleiotropic QTL and linkage disequilibrium (LD) between non-pleiotropic sites. To simulate these traits, 1675 SNP genotypes, already in LD from the founder population (average LD of 0.15 in this population), were simulated to serve as quantitative trait loci (QTL). At each population replicate, a random subset of 75 SNPs were assigned as pleiotropic QTL across all five traits, five random subsets of 25 SNPs each were assigned as pleiotropic QTL across four traits, ten random subsets of 50 SNPs each were assigned as pleiotropic QTL across three traits, and ten random subsets of 90 SNPs each were assigned as pleiotropic QTL across two traits. The remaining 75 SNPs were finally split in five groups of 15, to be assigned as QTL exclusive to each one of the five traits. This distribution of the QTL per trait is presented in the Venn diagram in Figure 1.

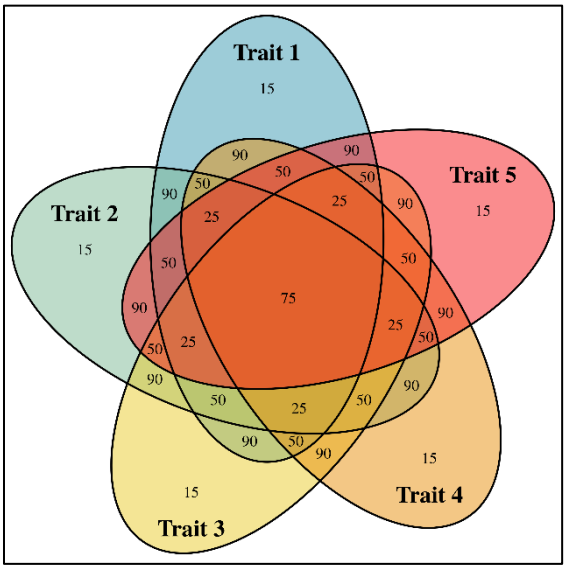

Figure 1. Venn Diagram describing the number of QTL shared among the five simulated traits.

Finally, QTL-effects were simulated, correlated between traits, so that the breeding values at the founder population presented

genetic correlations matching those obtained for the traits studied in the French Holstein dairy cattle population. The additive genetic variances of the simulated breeding values were set to $10 \times h^2$, with the heritabilities being those obtained for the real traits evaluated from the French Holstein dairy cattle data.

Individual-specific genetic correlations and their heritability estimates

For both the real and simulated data, *individual-specific genetic correlations* (iSGC) were calculated for sires, in order to evaluate how much differences in genetic correlations were expressed by different sires.

For the real data, iSGCs were calculated for all pairs of the five traits, evaluated with the bi-variate genetic evaluations models given by equation (1). Following the proposed by Cuyabano et al. (2024), sires with more than 500 daughters evaluated were selected, so that reliable genetic correlations could be obtained at the individual level, based on the daughters' EBVs. A minor change was made to calculate the iSGC, compared to how it was done by Cuyabano et al. (2024), who obtained the iSGC per sire by correlating the EBVs from their daughters. Here, prior to calculating the correlations between the daughters' EBVs from different traits, half of the dam's EBVs were subtracted from their daughters, so that on average, the iSGC comprised only sire information. Thus, for each sire s and for any pair of traits 1 and 2, their $i\text{-}th$ daughter's breeding values were corrected as:

$$\hat{g}_{1i,s} = \hat{g}_{1i} - \frac{\hat{g}_{1\{dam\ of\ i\}}}{2}, \quad (4)$$

$$\hat{g}_{2i,s} = \hat{g}_{2i} - \frac{\hat{g}_{2\{dam\ of\ i\}}}{2}, \quad (5)$$

for every $i=1, \ldots, n_s$. Finally, for each sire s:

$$iSGC_s = \frac{\sum_{i=1}^{n_s}\left(\hat{g}_{1i,s} - \overline{g}_{1,s}\right)\left(\hat{g}_{2i,s} - \overline{g}_{2,s}\right)}{(n_s - 1)\hat{\sigma}_{g_{s1}}\hat{\sigma}_{g_{s2}}}, \quad (6)$$

such that $\overline{g}_{1,s} = \frac{\sum_{i=1}^{n_s}\hat{g}_{1i,s}}{n_s}$ and $\overline{g}_{2,s} = \frac{\sum_{i=1}^{n_s}\hat{g}_{2i,s}}{n_s}$ are the mean daughters' corrected EBVs, and their variance are $\hat{\sigma}^2_{g_{s1}} = \frac{\sum_{i=1}^{n_s}\left(\hat{g}_{1i,s} - \overline{g}_{1,s}\right)^2}{n_s - 1}$ and $\hat{\sigma}^2_{g_{s2}} = \frac{\sum_{i=1}^{n_s}\left(\hat{g}_{2i,s} - \overline{g}_{2,s}\right)^2}{n_s - 1}$.

For the simulated data, iSGCs were calculated for all pairs of the five traits, only for the selected sires in the simulation routine. Since the simulations provided genotypes and the true simulated QTL effects, instead of using daughters' information, for each sire s, 500 gametes were simulated, at which QTL effects were applied. Thus, for each sire s and for any pair of traits 1 and 2, the additive genetic values of the $i\text{-}th$ gamete was given by:

$$\gamma_{1i,s} = \sum_{j=1}^{1675} X_i \alpha_{1j}, \quad (7)$$

$$\gamma_{2i,s} = \sum_{j=1}^{1675} X_i \alpha_{2j}, \quad (8)$$

for every $i=1, \ldots, 500$, such that α_{1j}'s and α_{2j}'s are the QTL effects (set as zero if the $j\text{-}th$ SNP is not a QTL for each of the traits). Finally, for each simulated sire s:

$$iSGC_{s(\sim)} = \frac{\sum_{i=1}^{500}\left(\gamma_{1i,s} - \overline{\gamma}_{1,s}\right)\left(\gamma_{2i,s} - \overline{\gamma}_{2,s}\right)}{(499)\hat{\sigma}_{\gamma_{s1}}\hat{\sigma}_{\gamma_{s2}}}, \quad (9)$$

such that $\overline{\gamma}_{1,s} = \frac{\sum_{i=1}^{500}\gamma_{1i,s}}{500}$ and $\overline{\gamma}_{2,s} = \frac{\sum_{i=1}^{500}\gamma_{2i,s}}{500}$ are the mean additive genetic values of the gametes, and their variance are $\hat{\sigma}^2_{g_{s1}} = \frac{\sum_{i=1}^{500}\left(\gamma_{1i,s} - \overline{\gamma}_{1,s}\right)^2}{499}$ and $\hat{\sigma}^2_{g_{s2}} = \frac{\sum_{i=1}^{500}\left(\gamma_{2i,s} - \overline{\gamma}_{2,s}\right)^2}{499}$.

Heritability estimates were obtained for the iSGCs, by treating them as a phenotype in a variance component estimation routine, using the pedigree relationship matrix for both the real and the simulated data, tracing back four generations from the sires. For the simulated data, heritability estimates were also obtained using a genomic relationship matrix (VanRaden, 2008) built from the simulated SNP-genotypes. The following model was used to estimate variance components:

$$iSGC = 1_n\mu + g + e, \quad (10)$$

in which $iSGC$ is the vector of iSGCs obtained for the n sires, between any two traits; μ is the overall mean; $g \sim N(0, A\sigma^2_{g(iSGC)})$ is the vector of breeding values associated to the iSGC, A is the pedigree relationship matrix (replaced by

the genomic relationship matrix G, for the simulated data), and $\sigma^2_{g(iSGC)}$ is the additive genetic variance associated to the iSGC; and $e \sim N(0, I_n\sigma^2_{e(iSGC)})$ is the vector of random residuals, and $\sigma^2_{e(iSGC)}$ is the residual variance.

Variance components for the iSGC were estimated through the residual maximum likelihood (REML; Patterson & Thompson, 1971), using the REMLF90 module from the BLUPF90 family of programs (Misztal et al., 2018). Finally, heritabilities of the iSGC were given by:

$$\hat{h}^2 = \frac{\hat{\sigma}^2_{g(iSGC)}}{\hat{\sigma}^2_{g(iSGC)} + \hat{\sigma}^2_{e(iSGC)}}. \quad (11)$$

Results & Discussion

Genetic parameters on real data

Heritability and genetic correlation estimates were obtained from the genetic parameters of the bi-variate genetic evaluations, for every pair of the five traits studied from the French Holstein dairy cattle population, and their values are presented in Table 1. These values agreed with those used for the French national genetic evaluation, as expected, and also agreed with reported heritabilities and genetic correlations between these traits. Finally, these values presented in Table 1 were the ones used as parameters to generate the breeding values for the simulated data, with genetic variances equal to $10 \times h^2$.

Table 1: Estimated heritabilities (diagonal bold values) and genetic correlations (upper triangle of the table) between the five traits studied in the French Holstein dairy cattle population. Values in gray indicate an estimate that was not statistically different from zero (significance level of 0.05).

	MY	PY	MSPD	SCS	CR
MY	**0.22**	0.78	-0.06	-0.04	-0.15
PY	---	**0.38**	-0.07	-0.01	-0.20
MSPD	---	---	**0.24**	0.31	-0.04
SCS	---	---	---	**0.13**	-0.26
CR	---	---	---	---	**0.01**

Distribution of the individual-specific genetic correlations on real and simulated data

Figures 2-5 present the distributions, in the form of density curves, of the iSGC obtained between the five traits studied, both on real and simulated data, indicating that different sires did present different potential for genetic correlations, expressed through their offspring.

The mean iSGCs on the real data presented bigger differences from the estimated genetic correlations with the Gibbs sampler, presented in Table 1 and indicated with dots at the x-axes of the plots, than the mean iSGCs on the simulated data. This could be due to the fact that, on real data, iSGCs were obtained for a subset of sires that had at least 500 daughters evaluated, rather than for all sires, potentially indicating a different mean iSGC for these elite sires, with respect to the whole population.

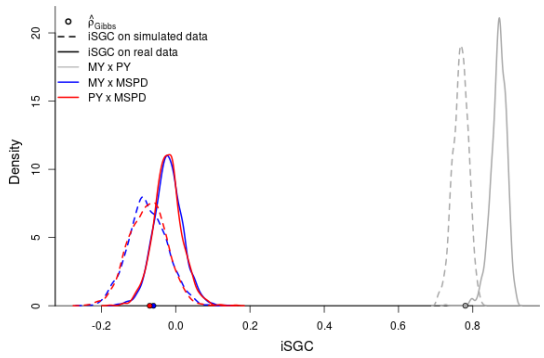

Figure 2. Distribution of the iSGC obtained across the pairs of the three production traits (MY, PY, and MSPD), on both real and simulated data.

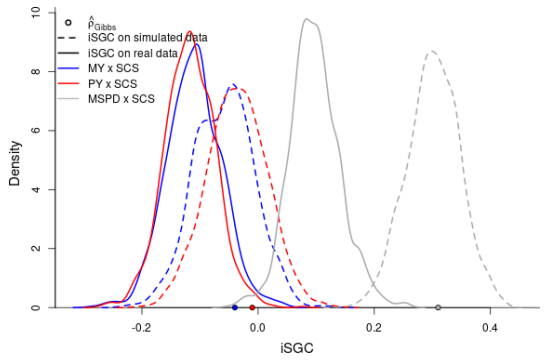

Figure 3. Distribution of the iSGC obtained between the production traits (MY, PY, and MSPD) and the health trait (SCS), on both real and simulated data.

Interestingly, on the real data, the overall iSGC between MY and CR and between PY and CR were less negative for the elite sires than the estimated genetic correlations between these traits, as shown in Figure 4. Conversely, the overall iSGC between MY and SCS and between PY and SCS were rather more negative (*i.e.* a stronger trade-off between these traits) for these elite sires than the estimated genetic correlations between these traits, as shown in Figure 3. If the hypothesis that iSGCs express a latent physiological trait holds, even if at least partially, these results suggest that selection is favoring a physiological trait that allows a better management of the trade-off between production and fertility, however in the detriment of the trade-off between production and health indicators. Nonetheless, it is important to note that a strengthening of the trade-off between traits does not mean that the traits themselves are not achieving genetic progress.

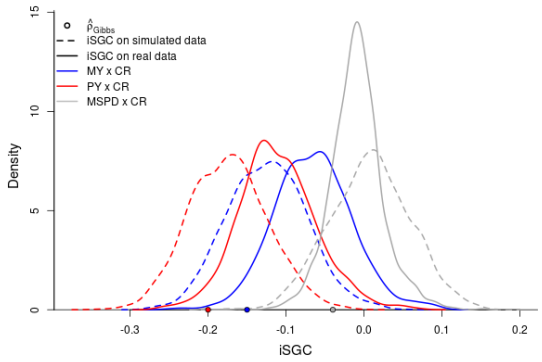

Figure 4. Distribution of the iSGC obtained between the production traits (MY, PY, and MSPD) and the fertility trait (CR), on both real and simulated data.

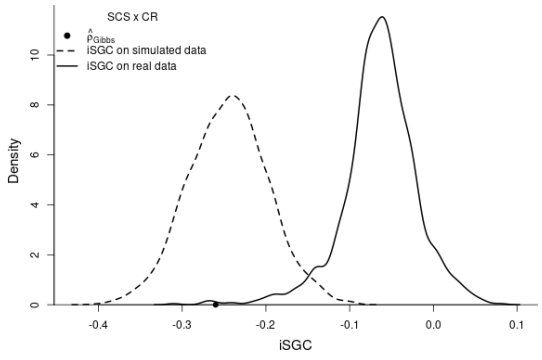

Figure 5. Distribution of the iSGC obtained between the health and fertility traits (SCS and CR), on both real and simulated data.

Heritabilities of the individual-specific genetic correlations on real and simulated data

Heritabilities were estimated for the iSGC, by treating them as a phenotype, as in the model presented in equation (10). These heritabilities were estimated for the iSGC obtained for both the real and simulated data. The goal of comparing these heritabilities of the iSGC on real data, to those of the iSGC on simulated data with the same genetic parameters, was to show that when no latent trait was associated to the differences between genetic correlations in a population, no heritabilities would be captured.

The estimated heritabilities are presented in Table 2, being the presented values for the heritabilities of iSGCs obtained on simulated data (lower triangle of Table 2), the obtained using the pedigree relationship matrix, since their values were not statistically different from the obtained with the genomic relationship matrix (significance level of 0.05). All these heritabilities of the iSGC on simulated data were not statistically different from zero (significance level of 0.05), indicating that neither family relationships, nor allele frequencies and LD patterns were enough to outline a genetic determinism for the different iSGC expressed by different sires.

With respect to the heritabilities of the iSGC on real data (upper triangle of Table 2), their values were significantly different from zero (significance level of 0.05), with the exception of the heritability of iSGC between MSPD and CR. Particularly, heritabilities of the iSGC between the two main production traits (MY and PY), between these main production traits and the health trait (SCS), and between these main production traits and the fertility trait (CR), were moderately high for dairy cattle traits, ranging from 0.38 to 0.51. These heritabilities suggest a reasonable level of genetic determinism associated to the different iSGC expressed by different sires, and these heritabilities could be due to the genetic

correlations at the individual level expressing, at least partially, a latent physiological trait.

Table 2: Heritability estimates for the iSGC obtained on the real data (upper triangle of the table), and for the iSGC obtained on the simulated data (lower triangle of the table). Values in gray indicate an estimate that was not statistically different from zero (significance level of 0.05).

	MY	PY	MSPD	SCS	CR
MY	---	0.45	0.16	0.45	0.46
PY	0.03	---	0.17	0.51	0.38
MSPD	0.02	0.02	---	0.23	0.05
SCS	0.02	0.02	0.02	---	0.11
CR	0.02	0.03	0.03	0.02	---

Conclusions

Genetic correlations, while treated as a parameter common to all individuals in genetic evaluations and selection indexes, may present different values across individuals in a population. By obtaining individual-specific genetic correlations for sires from a French Holstein dairy cattle population, this study has shown that indeed, different individuals present different patterns in their genetic correlations between five traits of interest. Moreover, individual-specific genetic correlations are heritable, suggesting that these parameters may be part of the expressions of a non-measurable (or latent) physiological trait. When it comes to traits that present a negative genetic correlation, the findings from this study may assist to select individuals better apt to manage the trade-off between traits. However, it remains a question of research, how to adequately and optimally use individual-specific genetic correlations and their heritability in a breeding program.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 Programme for Research & Innovation under grant agreement n°101000226.

References

Berry, D. P., Friggens, N. C., Lucy, M., & Roche, J. R. 2016. Milk Production and Fertility in Cattle. *Annual Review of Animal Biosciences*, *4*(1), 269–290.

Boichard, D., & Manfredi, E. 1994. Genetic Analysis of Conception Rate in French Holstein Cattle. *Acta Agriculturae Scandinavica, Section A - Animal Science*, *44*(3), 138–145.

Cuyabano, B. C. D., Croiseau, P., Shokor, F., Motta, M. R., Aguerre, S., & Mattalia, S. 2024. Genetic correlations: a parameter or a latent phenotype in genetic evaluations? *Interbull Bulletin*, *60*.

Hazel, L. N. 1943. The genetic basis for constructing selection. *Genetics*, *28*(6), 476–490.

Hazel, L. N., Dickerson, G. E., & Freeman, A. E. 1994. The Selection Index—Then, Now, and for the Future. *Journal of Dairy Science*, *77*(10), 3236–3251.

Miglior, F., Fleming, A., Malchiodi, F., Brito, L. F., Martin, P., & Baes, C. F. 2017. A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. *Journal of Dairy Science*, *100*(12), 10251–10271.

Misztal, I. S., Tsuruta, S., Lourenco, D. A. L., Masuda, Y., Aguilar, I., Legarra, A., & Vitezica, Z. 2018. *Manual for BLUPF90 family of programs*.

Patterson, H., & Thompson, R. 1971. Recovery of Inter-Block Information when Block Sizes Are Unequal. *Biometrika*, *58*(3), 545.

Pryce, J. E., Veerkamp, R. F., Thompson, R., Hill, W. G., & Simm, G. 1997. Genetic aspects of common health disorders and measures of fertility in Holstein Friesian dairy cattle. *Animal Science*, *65*(3), 353–360.

Rauw, W. M., Kanis, E., Noordhuizen-Stassen, E. N., & Grommers, F. J. 1998. Undesirable side effects of selection for high production efficiency in farm animals: a review. *Livestock Production Science*, *56*(1), 15–33.

Roxström, A., Strandberg, E., Berglund, B., Emanuelson, U., & Philipsson, J. 2001. Genetic and Environmental Correlations Among Female Fertility Traits and Milk Production in Different Parities of Swedish Red and White Dairy Cattle. A*cta Agriculturae Scandinavica, Section A - Animal Science*, *51*(1), 7–14.

VanRaden, P. M. 2008. Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, *91*(11), 4414–4423.

Windig, J. J., Calus, M. P. L., Beerda, B., & Veerkamp, R. F. 2006. Genetic Correlations Between Milk Production and Health and Fertility Depending on Herd Environment. *Journal of Dairy Science*, *89*(5), 1765–1775.