Genetic correlations and their dependence on environmental similarity
– insights from livestock data

Mads F. Schou a,b, Ary A. Hoffmann c & Torsten N. Kristensen b

a Department of Chemistry and Bioscience, Aalborg University, DK-9220 Aalborg East, Denmark
b Department of Biology, Lund University, Lund, SE-22362, Sweden
c School of BioSciences, Bio21 Institute, The University of Melbourne, Melbourne, VIC, Australia

Article type: Brief communication

Corresponding author: Mads Fristrup Schou, Department of Biology, Lund University, Lund, SE-22362, Sweden. Tel: +4551349299, e-mail: mads.schou@biol.lu.se.

Running head: Genetic correlations and environmental similarity

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/evo.13762.

This article is protected by copyright. All rights reserved.
Authors’ contributions: MFS, AAH and TNK conceived of the study; MFS and AAH collected data and carried out the statistical analyses; MFS, AAH and TNK drafted the manuscript.

Acknowledgements: This work was supported by the Danish Natural Research Council (DFF-8021-00014B to TNK).

Data archival: Data for this study are available in the supplementary information.

Abstract

Genetic correlations for a trait across environments are predicted to decrease as environments diverge. However, estimates of genetic correlations from natural populations are typically defined across a limited environmental range and prone to very large standard errors, making it difficult to test this prediction. We address the importance of environmental distance on genetic correlations by employing data from domestic cattle where abundant and accurate estimates are available from a wide range of environments. Three production traits related to milk yield show a clear decrease in genetic correlations with increasing environmental divergence. This pattern was also evident for growth traits and other yield traits but not for traits related to reproduction, morphology, physiology or disease. We suspect that this reflects weaker selection on these latter trait classes compared to production traits, or alternatively the effects of selection are constrained by unfavourable genetic correlations between traits. The results support the notion that traits that historically have been under strong directional selection in a small range of frequently encountered environments will evolve high genetic correlations across these environments, while exposure to uncommon (and dissimilar) environments lead to a reranking of gene effects and a decrease in genetic correlations across environments.

Key-words

Genetic correlations, environmental distance, genotype-by-environment interactions, cattle, selection.
Introduction

Genetic correlations for the same trait between environments are expected to decrease (corresponding to an increase in the genotype-environment interaction, $V_{ExG}$) as environments become less similar (Falconer and Mackay 1996; Sgrò and Hoffmann 2004). The core idea is that some genes will affect a trait only in specific environments while other genes are important for the trait across environmental conditions (Barton and Turelli 1989; Hoffmann and Parsons 1991). As environments diverge, changes in gene expression, and the contribution of a new set of environment-specific genes, cause a change in gene effects on a trait and thus a reduced genetic correlation (de Jong 1995; Falconer and Mackay 1996). Genetic correlations between environments can also evolve, as selection in different environments changes the underlying components of the genetic correlation, by altering the covariance between environments and the additive genetic variances in each environment. Understanding the change in genetic correlations across different environments and how such genetic correlations evolve over time is important in evolutionary biology where genetic correlations are typically considered to be constant when predicting and understanding evolutionary changes (Roff and Mousseau 1987; Via 1987; Pigliucci 2005; Arnold et al. 2008).

The extent to which genetic correlations evolve due to changes in the additive genetic variances and covariances may depend on the nature of the traits under consideration as well as the strength and mode of selection (Via and Lande 1985; Barton and Turelli 1989; de Jong 1990; Gromko 1995; Sgrò and Hoffmann 2004). Some traits are controlled by genes whose effects are not altered much by the environment, yet other traits might be governed by genes having highly environment-specific gene effects (Via and Lande 1985; de Jong 1990). While morphological traits determined early in development might not be influenced much by environmental conditions throughout much of an organism's life, some life history traits like reproduction are expected to be prone to environment-specific gene effects and might even show trade-offs: what increases fitness in one environment might decrease it in another (Roff 1992, 1996; Sgrò and Hoffmann 2004). Physiological traits might also be particularly prone to environment-specific gene effects because genes affecting them may only be expressed under certain environmental conditions, such as in the extreme case of high or low temperatures triggering genes controlling a quiescent phase, which then alters a whole range of physiological traits including metabolic rate and stress tolerance. If a trait is under stabilizing or directional selection, as is expected for most fitness-related traits, and the phenotypic optimum is stable across the typical environmental range, we expect genetic correlations within this environmental range to be positive and high. This is due to selection for stable genetic effects (large positive covariance across environments) and a resulting depletion of genetic variance (Via and Lande 1985, 1987). This selection is strongest in the most commonly encountered environments, and the impact of selection in environments encountered less frequently will typically be weak because of the rare
exposure to these environmental conditions (Via and Lande 1987). As a result, selection is expected to result in genetic correlations that are low between the typical environment and an environment outside the range of recent history, and more so when the two environments diverge.

Estimates of genetic correlations with small standard errors require large sample sizes because correlations represent the ratio of the covariances across environments over the additive genetic variances, all measured with error (Falconer and Mackay 1996). High quality data with sufficient sample sizes are hard to obtain from natural populations and therefore these predictions are difficult to test. In contrast, livestock studies, and especially data sampled from domestic cattle, can include thousands of animals tested in a range of environments, resulting in far more accurate estimates of genetic correlations across environments. Farm animals or cultivated plants are typically selected to perform well across a rather limited range of environments aided by e.g. homogenous rearing and feeding conditions across wide geographical ranges. This is expected to lead to strong positive genetic correlations across environments unless they fall outside the normal range encountered (Boettcher et al. 2003; Kearney et al. 2004; Haile-Mariam et al. 2008; Strandberg et al. 2009). In domestic farm animals, selection for maintaining strong positive genetic correlations across environments may be particularly intense for traits related to yield and product quality, as genes yielding consistent performance within typically encountered environments are desirable, but less so for other traits (Rauw et al. 1998; Simm 1998; Hill 2016).

Here we use data from domestic cattle to test two predictions. First, genetic correlations within traits are predicted to become weaker as environments diverge, and second, this trend is expected to be stronger for traits closely associated with fitness. Our results show that production traits exhibit a significant decrease in genetic correlations as environments become less similar, whilst other trait categories do not show this pattern. This dependency of the nature of the trait categories likely reflects distinct selection operating on different traits.

Materials and methods

The survey

Studies of livestock partitioning the phenotypic variance across environments typically estimate genotype by environment interactions by calculating the genetic correlation of a trait between two environments (Falconer 1990). We undertook a literature search using the search term “environments” AND “genetic correlation*” AND “cattle” (assessed: 15/09/17) on the Web of Science Core Selection (www.webofknowledge.com). This search produced 377 articles, from which we
only accepted articles where: 1) genetic correlations were calculated within traits across at least two environments, 2) mean trait values were given for each environment, enabling estimation of differences between environments, and 3) the environments were clearly defined, e.g. by geography, abiotic conditions or rearing conditions. Studies where the environments were solely defined by differences in mean trait values among the environments were omitted. The reference lists of the identified papers were used to identify additional papers, complying with the above requirements. Only papers written in English were included.

From the papers complying with the above criteria we extracted estimates of genetic correlations (and standard errors if available), trait means, sample sizes (number of individuals and/or records) and environmental details. When estimates were only available graphically, we extracted numerical data using WebPlotDigitizer (http://arohatgi.info/WebPlotDigitizer/app/). Some studies used random regression analysis to estimate genetic correlations across a defined environmental dimension such as ambient temperature (e.g. Ravagnolo et al. 2000). For these studies, we extracted estimates of genetic correlations at the most extreme environments and the more optimal environments. In several studies, multiple environments were considered, whereby the genetic correlations from these comparisons are non-independent. For instance, for environments A, B and C, usage of all reported genetic correlations (A-B, A-C and B-C) will cause nested non-independence within the study as each environment and the individuals measured therein, is used for two comparisons. These estimates are however highly informative as they represent different environmental distances tested with many individuals of shared and estimable ancestry (i.e. common sires) across several environments, contributing to an accurate estimate of differences in the additive genetic covariance and genetic correlation. The non-independence can partly be dealt with by having study as a random effect in our statistical models. However, to be conservative and reduce the number of comparisons while retaining the full range of environments, we ordered environments according to phenotypic mean and then removed all contrasts not including the environment with the highest phenotypic mean (i.e. if A has the highest mean, only A-B and A-C were retained). In this way, the random effect of study captures both study and environment A as a baseline. In total, our approach yielded 310 genetic correlation estimates from 49 studies (Table S1).

Trait definitions

Traits under direct farmer control such as age of culling, and traits that could not be related to a distinct biological trait, such as lifetime net income, were excluded (see Table S2 for further details on trait definitions). When possible, weaning weight was used instead of birthweight as a measure of growth. Weaning weight is more informative
due to the larger experience of the calf in the reported environment and the maternal effect on this trait is presumably smaller than on birthweight. If a paper provided data on several lactations, we used data from the first lactation, as this was the most prevalent across studies, and as this reduced issues associated with e.g. culling of low production individuals (Banos and Shook 1990). Trait estimates related to milk production were typically reported on Holstein cattle (other names used in the literature: Holstein-Friesian, Friesian and Holstein), while trait estimates related to beef production were mostly reported from Angus and Hereford or various mixed breeds. In order to test if genetic correlations have been affected by selection across traits of shared evolutionary history, each trait was classified into one of five categories based on evolutionary principles: disease indicator, growth, life history, morphology and physiology (Roff and Mousseau 1987; Hoffmann et al. 2016) (Table S1). When reviewing our final dataset, we observed a large proportion of milk production traits in the trait category “life history”, which typically had large sample sizes and was measured across a wide range of environments (measured by trait mean) enabling trait-specific analyses. We therefore split the life history category into a group of female reproductive traits and three milk production trait-groups for further analyses.

Statistical analyses

We undertook two different approaches to quantify the environmental difference underlying the genetic correlations. In the first approach we quantified the difference in trait mean between two environments using the response ratio (lnRR) (Nakagawa et al. 2015)

\[
\ln RR = \ln \frac{\mu_1}{\mu_2} = \ln \frac{\text{trait mean}_1}{\text{trait mean}_2}
\]

lnRR is by definition always positive, and a larger value represents a larger difference in trait mean between the two environments. We modelled the change in the genetic correlation with lnRR using linear mixed models (Gaussian distribution) in the lme4 package (v. 1.1-17, Bates et al. 2015) in R (R Core Team 2019). The models contained lnRR as the single fixed effect and we accounted for the effect of study and trait, and if applicable breed, by including these as random effects in the models. The statistical significance of lnRR was assessed by comparing the full model with a null model without the lnRR term using a likelihood ratio test. The models were fitted using REML but refitted with ML when performing the likelihood ratio tests. The distribution of our response variable, the genetic correlation, behaved as a proportion with an upper bound of 1 thus potentially violating the assumption of
homogeneity of variance. However, genetic correlations can also be negative, indicating a change in sign in the effects of some genes (two estimates in our dataset) whereby treating genetic correlations as a proportion has little biological meaning. We chose to analyse and present the raw data, but also test the robustness of the models by repeating our analyses with an arcsine transformation while restricting the parameter space to 0;1. This additional analysis confirmed the findings of the initial analysis (patterns consistent with those reported below) and results are available as supplementary material (Figures S1 and S2). Only 44% of the studies reported genetic correlations with associated standard errors, preventing us from implementing an analysis of sampling variance in our analysis. However, in two of the analysed milk production trait-groups more than 50% of the estimates had affiliated standard errors (milk yield and protein yield), and we therefore redid the analyses of these traits while considering the standard errors to confirm our findings. The R-package MCMCglmm (v.2.25) (Hadfield 2010) has a built-in argument (mev) specifically designed to handle sampling variance of data points in linear mixed models and we therefore proceeded with this modelling approach. As above we specified lnRR as the sole fixed effect as well as study and trait as a random effect, but we also included a term (mev) specifying the sampling variance (standard error squared) for each data point. We specified the random effects and residual priors as (uninformative) inverse gamma priors (V = 1, nu = 0.002). Convergence of the estimates was checked by inspecting trace plots of the MCMC chain and the level of autocorrelation among posterior samples.

For the second approach we assigned geographical regions into tropical, arid or temperate-cold according to the Köppen-Geiger climate classification (Peel et al. 2007) (Table S1, see Figure S3 and S4 for change in trait mean in these climate classifications). The purpose of this alternative approach was to develop a measure of environmental differences that was independent of performance (trait mean). Hence this measure only captures macro-differences in climate between environments and does not include differences in environment that is unrelated to the Köppen-Geiger climate classification (i.e. rearing conditions) that are likely to affect performance. Genetic correlations were split according to this classification system into those within one climate and those between two different climates. In order to test for an effect of climate similarity on genetic correlations, we followed the approach described above and constructed linear mixed models for each trait/trait category. As above, the effects of study, breed and trait were modelled as random effects and the sole fixed effect was the two-level factor climate (similar or different).

Results and discussion

For commercial dairy cattle production, the most important performance trait is milk yield. This trait, and the other yield traits, showed decreasing genetic correlations as environments became less similar (Figure 1). This was robust to the implementation
of reported standard errors when applied to the traits where this was possible (protein yield: pMCMC = 0.018; milk yield: pMCMC = 0.003). The statistical support for this was strongest when trait means were used as a measure of environmental similarity. Milk-yield and fat-yield also showed a significant decrease in the genetic correlations across different climates compared to similar climates, while protein yield showed a trend in this direction (Figure 1). In summary, these results suggest that there are environment-specific gene effects for yield traits in dairy cattle increasing in impact as environments become less similar.

There is evidence that domestication in cattle occurred approximately 10,500 years ago (Bollongino et al. 2012; Scheu et al. 2015). Artificial selection for increased production in cattle has likely taken place ever since. However the efficiency of artificial selection has been markedly improved lately, aided by developments including artificial insemination, identification of superior bulls through progeny performance, use of analytical methods such as Best Linear Unbiased Prediction (BLUP), and finally by genomic selection (Simm 1998; Dobson et al. 2007; Kristensen et al. 2015; Hill 2016). In dairy cattle, the combined effect of artificial selection and improved feeding and housing conditions, have led to marked increases in milk yield during the 20th and 21st century (Simm 1998; Hill 2016).

Most of the recent selection has taken place in temperate environments, in particular in the Holstein and Jersey breeds, and genetic material has been exported worldwide. Our findings show that the strong directional selection for performance (here milk yield) in the most common environments has led to maintenance of high genetic correlations across the narrow range of these environments. However, selection in the optimal environments has also shifted the genetic composition of animals so that a substantial re-ranking of gene effects causes a reduction in the genetic correlations as we go to different environments, rarely encountered during the selection process. Thus, the world-wide use of semen from a few elite sires, typically from Western Europe or North America, and export of dairy cattle from e.g. Europe to Africa does constitute a problem. Our results should encourage focussing on reducing $V_{ExG}$ and / or using locally adapted breeds / varieties in plant and animal breeding.

For traits related to growth, which for beef cattle is the primary production trait, we found a similar pattern. Despite the low proportional change in mean value for growth traits compared to milk yield, this data showed a significant decrease in genetic correlations with increasing environmental difference. This could not be confirmed when we used climate to predict the relative size of the genetic correlation (Figure 2). Traits related to meat production have likely been under artificial selection for as long as cattle have been domesticated. Thus, as in the case of milk production traits, our results are in accordance with this history of selection, although the signature is weaker. One reason for a weaker signature might be less intensive selection in beef cattle; dairy cattle breeding is at the forefront when it comes to applying new technologies and thereby making genetic progress for desired traits (Simm 1998).
For the trait categories morphology, physiology, reproductive performance and disease indicators, we did not find evidence of a decrease in genetic correlations when environments diverged, neither when defined by trait means nor by climatic zones. The lack of environmental effects does not reflect power issues as data available for comparison was similar to that available for milk yield traits, and the range of environments was also similar (Figures 1 and 2). There are several possible explanations for this contrasting pattern. One possibility is that there are fewer environment-specific gene effects for these types of traits. This is in accordance with the expectations for morphological traits, but seems unlikely in particular in the case of female reproductive performance (fitness related traits) where environment-specific gene expression is pronounced (Roff 1992, 1996; Sgrò and Hoffmann 2004). Another possibility is that selection on these traits is weaker than on production traits which have experienced a long history of directional selection (Simm 1998). Under relaxed selection on traits, there is weaker selection for gene effects that affect a trait in the same way in typically encountered environments, but this does not explain the overall high genetic correlations. Finally, selection on the different trait categories is not necessarily independent. For instance some female reproduction traits and disease indicator traits are genetically and phenotypically correlated with production traits – often in an unfavourable direction (Rauw et al. 1998). This means that the response to selection on female reproduction traits and disease indicator traits in the commonly experienced environment might be slow, but this does again not provide a satisfactory explanation for why genetic correlations are high and remain high across the environmental range.

Genetic correlations remain hard to estimate in natural populations (Sgrò and Hoffmann 2004; Akesson et al. 2008) and the few estimates available in the literature often show a shift in the sign of the genetic correlations across environments (Sgrò and Hoffmann 2004; Stinchcombe et al. 2010). For instance, a well-executed meta-analysis of laboratory derived estimates, suffered from low power and taxonomic biases limiting inference on the effect of environmental differences on the genetic correlations between traits (Wood and Brodie 2015). Our findings suggest strong genetic correlations across similar environments and a decrease in genetic correlations with increasing environmental distance for traits under strong directional selection. Whether these findings apply to natural populations, where the strength and direction of selection may not be as consistent and heritability estimates typically are lower, remains to be tested, and we advocate for more studies from natural or semi-natural populations. Such studies have been hampered by low sample sizes, but we expect that the continued development of sequencing techniques and the decrease in costs will allow for low-cost establishment of large-scale pedigrees of natural or semi-natural populations and bring us new insights on this subject.

This article is protected by copyright. All rights reserved.
References


Falconer, D. S. 1990. Selection in different environments: effects on environmental sensitivity (reaction norm) and on mean performance. Genet. Res. 56:57–70.


This article is protected by copyright. All rights reserved.


Figures and figure captions

Figure 1. The association between genetic correlations and the environmental difference in life history traits associated with milk production in dairy cattle. A genetic correlation is calculated from the performance of related individuals in two distinct environments. We used two different approaches to quantify the difference between these two environments; we estimated the response ratio in trait mean (lnRR, difference in trait mean) between the two environments and we used Köppen-Geiger climate classification to group geographical regions into tropical, arid or temperate-cold climates such that genetic correlations could be classified as within one climate (similar) or between two climates (different). For traits with a significant effect of
InRR, we plotted the predicted association from the model, along with 95% confidence intervals obtained by parametric bootstrapping (n = 10,000). *P < 0.05.
Figure 2. The association between genetic correlation and environmental difference in five evolutionary trait categories for dairy and beef cattle. A genetic correlation is calculated from the performance of related individuals in two distinct environments. We used two different approaches to quantify the difference between these two environments; we estimated the response ratio in trait mean (lnRR, difference in trait mean) and we used Köppen-Geiger climate classification to group geographical regions into tropical, arid or temperate-cold climates such that genetic correlations could be classified as within one climate (similar) or between two climates (different). From the life-history category only traits related to reproductive performance in females are used as the remaining traits were analysed separately (Figure 1). When analysing the growth category, we omitted the trait daily weight gain (n = 2) as this trait was highly influential and caused overconfidence with the observed pattern (P < 0.001). For traits with a significant effect of lnRR, we plotted the predicted association from the model, along with 95% confidence intervals obtained by parametric bootstrapping (n = 10,000).
Author/s: Schou, MF; Hoffmann, AA; Kristensen, TN

Title: Genetic correlations and their dependence on environmental similarity-Insights from livestock data

Date: 2019-08-01


Persistent Link: http://hdl.handle.net/11343/286017