Covariance reaction norms: A flexible approach to estimating continuous environmental effects on quantitative genetic and phenotypic (co)variances

Jordan S. Martin

Human Ecology Group, Institute of Evolutionary Medicine,
University of Zurich, Switzerland

jordan.martin@uzh.ch

Abstract

Estimating quantitative genetic and phenotypic (co)variances plays a crucial role in investigating key evolutionary ecological phenomena, such as developmental integration, life history tradeoffs, and niche specialization, as well as in describing selection and predicting multivariate evolution in the wild. While most studies assume (co)variances are fixed over short timescales, environmental heterogeneity can rapidly modify the variation of and associations among functional traits. Here I introduce a covariance reaction norm (CRN) model designed to address the challenge of detecting how trait (co)variances respond to continuous environmental change, building on the animal model used for quantitative genetic analysis in the wild. CRNs predict (co)variances as a function of continuous and/or discrete environmental factors, using the same multilevel modeling approach taken to prediction of trait means in standard analyses. After formally introducing the CRN model, I validate its implementation in Stan, demonstrating unbiased Bayesian inference. I then illustrate its application using long-term data on cooperation in meerkats (*Suricata suricatta*), finding that genetic (co)variances between social behaviors change as a function of group size, as well as in response to age, sex, and dominance status. Accompanying R code and a tutorial are provided to aid empiricists in applying CRN models to their own datasets.

Keywords: GxE, plasticity, flexibility, multivariate, mixed effects, animal model

Introduction

Accurately estimating phenotypic and quantitative genetic (co)variances is essential for understanding multivariate evolution in the wild. For instance, quantifying the (co)variances of thermoregulatory traits and growth rates is crucial for explaining differential patterns of population adaptation and divergence in response to climate change (de la Mata et al., 2022; Oomen & Hutchings, 2022; Schaum et al., 2022). Empirical estimates of covariance between life history traits are also critical for testing theoretical models of putative tradeoffs (negative covariances) between growth, maintenance, survival, or reproduction (Haave-Audet et al., 2022; Chang et al., 2023), which are hypothesized to constrain the direction and rate of adaptive evolution (Stearns, 1989; Roff, 1996). Positive genetic covariances may instead accelerate adaptation across environments, such as in red flour beetles (*Tribolium castaneum*), where selection for drought resistance has been found to indirectly select for greater heat resistance via a correlated genetic response (Koch et al., 2020). Estimating phenotypic (co)variances is similarly important for addressing various challenges in evolutionary ecology, such as distinguishing between repeatable and stochastic patterns of trait selection in the wild (Dingemanse et al., 2021; Martin, 2021), testing theoretical models of developmental integration and niche specialization (Damián et al., 2020; Rolian, 2020; Martin et al., 2023), as well as for making evolutionary predictions in systems undergoing rapid environmental change or exhibiting processes of non-genetic inheritance, such as cultural learning and niche construction (Danchin & Wagner, 2010; Fogarty & Wade, 2022).

For polygenic and environmentally responsive traits, the quantitative genetic **G** matrix and phenotypic **P** matrix can be used to describe these multivariate (co)variances and predict their evolutionary consequences (Lande, 1979; Lande & Arnold, 1983). Various quantities derived from **G** and **P** matrices have also long been of interest in evolutionary genetics and ecology, such as covariance tensors and principal components (Schluter, 1996; Aguirre et al., 2014) for comparing divergence across populations (McGlothlin et al., 2018; Royauté et al., 2020), or canonical axes (Phillips & Arnold, 1989; Blows & Brooks, 2003) for describing (non)linear selection on correlated phenotypes (Brooks et al., 2005; Oh & Shaw, 2013). Multivariate, multilevel regression models (also known as mixed effects, hierarchical, or random regression models) are well-established in the literature and widely applied for empirically estimating **G** and **P** matrices (e.g. Nussey et al., 2007; Dingemanse & Dochtermann, 2013; Brommer et al., 2019). Multivariate animal models—a specific form of generalized multilevel regression model—are particularly useful for quantitative genetic analysis, as they can take full advantage of naturally occurring, continuous variation in genetic relatedness and environmental conditions across subjects (Kruuk, 2004; Wilson et al., 2010). This allows the animal model to provide greater flexibility and robustness for describing heritable (co)variation in wild populations, in comparison to classical methods that rely on the assumptions of balanced breeding experiments or specific kin-class comparisons (Kruuk & Hadfield, 2007). Building on the well-established animal model, the present paper develops flexible extensions for predicting variation in **G** and **P** matrices attributable to continuous environmental effects.

Motivation for a novel method

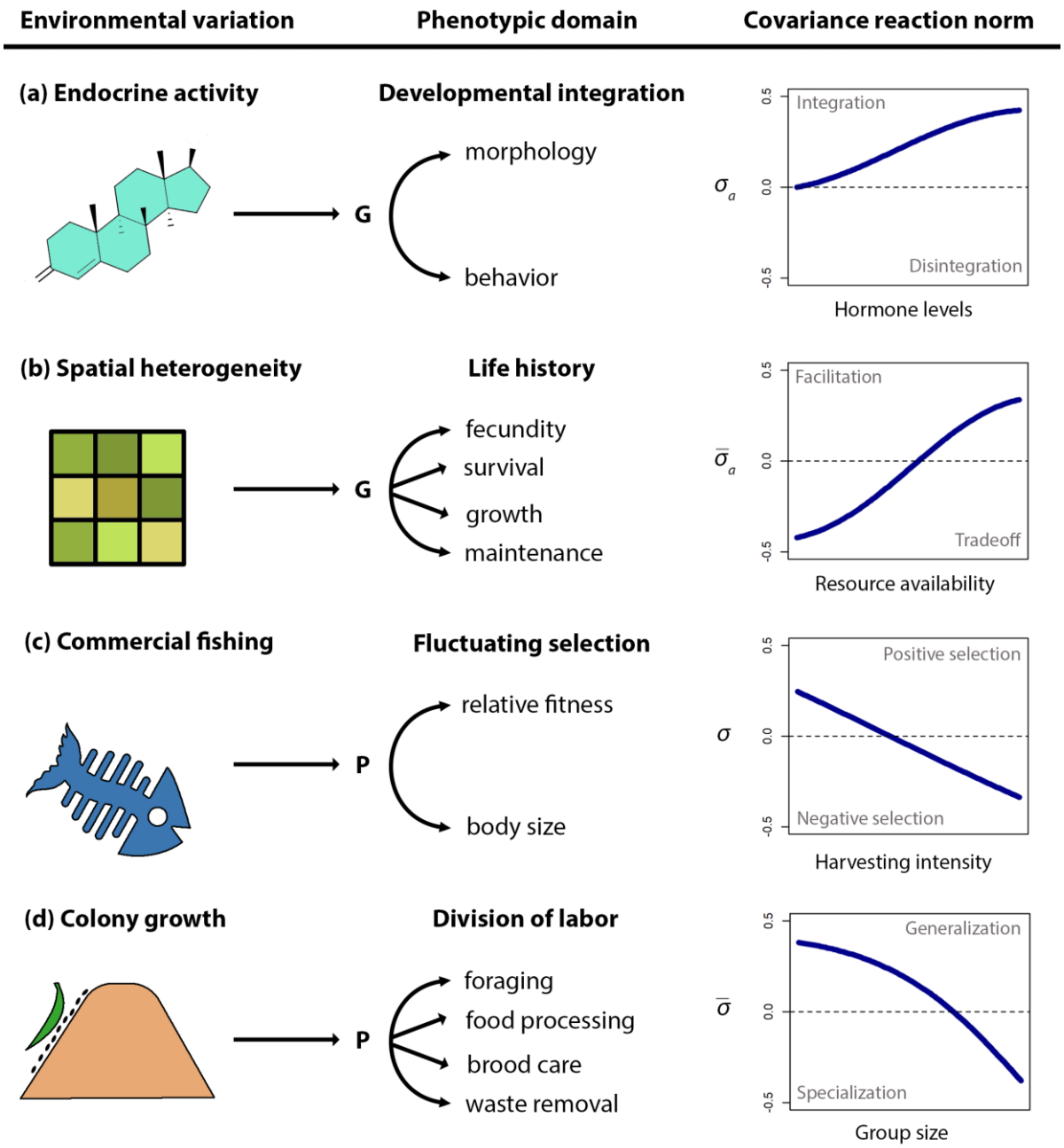
Despite longstanding theoretical interest and empirical evidence of both micro- and macroevolutionary stability in **G** and **P** matrices (Björklund, 1996; Estes & Arnold, 2007; Henry & Stinchcombe, 2023; McGlothlin et al., 2018), genetic and phenotypic (co)variances are also expected to change rapidly within many taxa across space and time, as individuals face continuously varying environmental conditions that predictably shape the expression and selection of their functional traits (**Fig. 1**). For example, previous research across a wide range of taxa (e.g. lizards, Yewers et al., 2017; Wittman et al., 2021; flies, Carvalho & Mirth, 2015; frogs, Lofeu et al., 2017; mice, vom Saal, 1979;

Huber et al., 2017; and primates, Montoya et al., 2013; Grebe et al., 2019) has shown that endocrine activity and the resulting hormonal milieu experienced during both prenatal and postnatal development exhibit dose-dependent effects on the integration (positive genetic covariance) of various morphological and behavioral phenotypes in adult organisms (**Fig. 1a**). As another example, consider that classic theoretical models (van Noordwijk & de Jong, 1986) predict associations among life history traits to be contingent on the relative importance of among-individual differences in resource acquisition versus allocation. As a consequence, spatial or temporal heterogeneity in factors such as resource availability are expected to cause continuous variation in the genetic effects acting to constrain (negative genetic covariance, i.e. tradeoffs) or facilitate (positive genetic covariance) ongoing adaptation (Mats Björklund, 2004; Mats Björklund & Gustafsson, 2015; Haave-Audet et al., 2022); **Fig. 1b**). Similarly, continuous fluctuations in selection are expected to occur when the fitness effects of quantitative traits vary across functional contexts, as described by changes in the covariance between relative fitness and phenotype (Lande, 1976). In many fish, for instance, large body size reduces predation risk and promotes greater mating and reproductive success (Barneche et al., 2018; Uusi-Heikkilä, 2020); however, commercial harvesting of fish also tends to target larger individuals (Sharpe & Hendry, 2009; Heino et al., 2015), facilitating continuous shifts in the strength and direction of selection on size as a function of the intensity of local harvesting (**Fig. 1c**), likely in interaction with other factors such as local predator densities and sex ratios (Uusi-Heikkilä, 2020; Jusufovski & Kuparinen, 2020). Finally, both theory (Bonner, 2004; Jeanson et al., 2007) and extensive empirical study (e.g. Karsai & Wenzel, 1998; Thomas & Elgar, 2003; Ferguson-Gow et al., 2014; Ulrich et al., 2018) have demonstrated that division of labor can emerge spontaneously during colony growth in eusocial species, with workers exhibiting generalist phenotypes at small group sizes (average positive phenotypic covariance among tasks) but shifting toward specialist phenotypes as group size increases (negative phenotypic covariance; **Fig. 1d**).

These dynamic patterns of genotype-, phenotype-, and fitness-by-environment interaction can be formally quantified by changes in **P** and **G** matrices across contexts. Current multivariate animal models are well suited for characterizing discrete changes in trait (co)variances due to categorical environmental effects, such as experimental conditions (e.g. solitary versus group housing) and developmental stages (e.g. juvenile versus adult) or subjectively binned environmental covariates from the field (e.g. high versus low quality habitats). This is typically achieved through a so-called character state approach, where separate models are fit for trait expression in each discrete environmental state and individuals' additive genetic (breeding) values are allowed to correlate across models (Via & Lande, 1985; Lynch & Walsh, 1998). However, as argued above, environmental effects on **P** and **G** matrices will often reflect continuous and potentially nonlinear processes that are challenging to describe with character-state models (**Fig. 1**). These complex dynamics can be interpolated post-hoc from estimates of discrete states (Mitchell & Houslay, 2021), but this strategy will often require prohibitively large sample sizes for accurate inference due to discretizing the problem into $s\frac{p(p-1)}{2}$ distinct and independently estimated covariance terms, where p is the number of phenotypes and s is the number of states necessary to effectively approximate the underlying function (which may be very large for complex responses to multivariate environments). As a consequence, reliance on interpolation through character-state models will tend to reduce statistical power for detecting complex functional relationships in heterogeneous environments. Fortunately, mathematically complementary reaction norm models (de Jong, 1995; Lynch & Walsh, 1998; Nussey et al., 2007) can be used to more directly and parsimoniously describe these processes, taking full advantage of available environmental information with much fewer parameters. The following section introduces such a 'covariance reaction norm' (CRN) approach to estimating continuous and potentially nonlinear environmental effects on trait (co)variances. I subsequently validate this model for empirical

application using simulation-based calibration (Talts et al., 2018) and explore its application through a worked empirical example on cooperative task specialization in meerkats (*Suricata suricatta*). Accompanying code and a guided tutorial for implementation of CRN models in the R statistical environment (R Core Team, 2023) using the Stan statistical programming language (Carpenter et al., 2017) can be found on Github (see **data availability**).

Figure 1. Potential empirical applications for covariance reaction norm models.



Footnote. Four examples (**a-d**) are shown of phenotypic domains (middle column) where continuous environmental variation (left column) is likely to cause continuous changes in quantitative genetic (**G**; top rows) and phenotypic (**P**; bottom rows) trait covariances, as formally described by hypothetical covariance reaction norms (CRNs; right column) quantifying patterns of continuous GxE and PxE across environmental states. See the main text for a detailed description of each scenario and **Eq. 2-3** for a formal description of how such CRNs can be empirically estimated.

Covariance reaction norms

Quantitative genetic analysis

The animal model allows for partitioning random quantitative genetic effects **G** and environmental effects on phenotypes. Extensive prior work has provided detailed overview of the animal model and its various extensions (e.g. Nussey et al., 2007; Wilson et al., 2010; Thomson et al., 2018; Martin & Jaeggi, 2022). Therefore, I focus herein on a highly simplified presentation of the animal model to highlight novel extensions, as well as to avoid detailed discussion of general issues in regression analysis such as the inclusion of various kinds of fixed and random effects. A multivariate animal model can be specified for each of *p* Gaussian phenotypes $\left[\boldsymbol{z_1^\top}, \ldots, \boldsymbol{z_p^\top}\right]^\top$ measured for *n* individuals by

$$\begin{bmatrix} \boldsymbol{z_1} \\ \vdots \\ \boldsymbol{z_p} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_1 + \boldsymbol{\alpha_1} + \boldsymbol{\epsilon_1} \\ \vdots \\ \boldsymbol{X\beta}_p + \boldsymbol{\alpha_p} + \boldsymbol{\epsilon_p} \end{bmatrix} \quad (1.1)$$

where **X** is an *n* x *b* matrix of *b* continuous and/or discrete covariates (e.g. local density, age, sex, resource abundance, seasonal precipitation and temperature, etc.), and $\left[\boldsymbol{\beta_1^\top}, \ldots, \boldsymbol{\beta_p^\top}\right]^\top$ are *b* x 1 vectors of trait-specific fixed effect sizes including global intercepts. After adjusting for these effects, the model estimates trait-specific additive genetic (i.e. breeding) values $\left[\boldsymbol{\alpha_1^\top}, \ldots, \boldsymbol{\alpha_p^\top}\right]^\top$ and residual environmental values $\left[\boldsymbol{\epsilon_1^\top}, \ldots, \boldsymbol{\epsilon_p^\top}\right]^\top$. Further genetic effects due to dominance or epistasis can also be parameterized when relevant for the goals of the analysis, along with any other random intercepts or slopes of interest. If repeated individual-level measurements are available, residuals can also be further partitioned into permanent and stochastic environmental components. Trait (co)variances due to additive genetic and residual effects are assumed to be approximated by multivariate normal distributions

$$\begin{bmatrix} \boldsymbol{a_1} \\ \vdots \\ \boldsymbol{a_p} \end{bmatrix} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{G} \otimes \boldsymbol{A}); \begin{bmatrix} \boldsymbol{\epsilon_1} \\ \vdots \\ \boldsymbol{\epsilon_p} \end{bmatrix} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\Sigma}) \quad (1.2)$$

With the **G** matrix being scaled using the Kronecker product $\otimes$ by a relatedness matrix **A** that quantifies pairwise relatedness among subjects, calculated using standard pedigree methods or molecular approaches. This basic animal model structure assumes that phenotypic (co)variances described by the **G** matrix are constant across subjects, adjusted for any other fixed and random effects predicting phenotypic means. The goal is now to relax this assumption by also allowing for fixed effects due to continuous or discrete environmental factors to also predict variation in trait (co)variances.

Predicting genetic (co)variances

The **G** matrix can be parameterized using genetic variances σ_a^2 and correlations r_a such that

$$\boldsymbol{G}: \begin{bmatrix} \sigma_{a_1}^2 & \cdots & r_{a_{1,p}}\sigma_{a_1}\sigma_{a_p} \\ & \ddots & \vdots \\ & & \sigma_{a_p}^2 \end{bmatrix} \quad (1.3)$$

where genetic covariances $\sigma_{a_{1,p}} = r_{a_{1,p}}\sigma_{a_1}\sigma_{a_p}$ are given by the product of genetic correlations and standard deviations (square roots of the genetic variances). Note that bold symbols are used to distinguish vectors and matrices from scalars. Separating out the scale of variation σ_a^2 for each variable from their standardized associations r_a is useful for further expanding the model, as environmental

factors may exhibit independent effects on these distinct components, which would otherwise be confounded together through direct prediction of the covariance. Using Eq. 1.3, the basic animal model can now be expanded to a covariance reaction norm (CRN) model by using link functions to predict how genetic variances and correlations change in response to the same matrix **X** of environmental covariates used to predict phenotypic means (or a relevant subset of these predictors). In particular, using the subscript (X_n) to denote the **G** matrix predicted from a CRN in the environmental context measured for subject *n*

$$\begin{bmatrix} \boldsymbol{z_1} \\ \vdots \\ \boldsymbol{z_p} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_1 + \boldsymbol{\alpha}_{(X)_1} + \boldsymbol{\epsilon_1} \\ \vdots \\ \boldsymbol{X\beta}_p + \boldsymbol{\alpha}_{(X)_p} + \boldsymbol{\epsilon_p} \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} \boldsymbol{a}_{(X)_1} \\ \vdots \\ \boldsymbol{a}_{(X)_p} \end{bmatrix} \sim \boldsymbol{N}\left(\boldsymbol{0}, \boldsymbol{G}_{(X)} \otimes \boldsymbol{A}\right);\ \boldsymbol{G}_{(X_n)}: \begin{bmatrix} \sigma^2_{a(X_n)_1} & \cdots & r_{a(X_n)_{1,p}}\sigma_{a(X_n)_1}\sigma_{a(X_n)_p} \\ & \ddots & \vdots \\ & & \sigma^2_{a(X_n)_p} \end{bmatrix}$$

$$\begin{bmatrix} \log\left(\boldsymbol{\sigma^2_{a(X)_1}}\right) \\ \vdots \\ \log\left(\boldsymbol{\sigma^2_{a(X)_p}}\right) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_{\sigma^2_1} \\ \vdots \\ \boldsymbol{X\beta}_{\sigma^2_p} \end{bmatrix};\quad \begin{bmatrix} \operatorname{atanh}\left(\boldsymbol{r_{a(X)_{1,2}}}\right) \\ \vdots \\ \operatorname{atanh}\left(\boldsymbol{r_{a(X)_{p-1,p}}}\right) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_{r_1} \\ \vdots \\ \boldsymbol{X\beta}_{r_{p-1,p}} \end{bmatrix}$$

Rather than defining a single genetic variance and set of correlations for each response variable, as in the standard animal model (Eq. 1), the CRN animal model predicts *n* **G** matrices $\boldsymbol{G}_{(X)} = \left(\boldsymbol{G}_{(X_1)}, \ldots, \boldsymbol{G}_{(X_n)}\right)$ each composed of context-specific genetic variances $\boldsymbol{\sigma^2_{a(X)_p}} = [\sigma^2_{a(X_1)_p}, \ldots, \sigma^2_{a(X_n)_p}]'$, and correlations $\boldsymbol{r_{a(X)_{1,p}}} = [r_{a(X_1)_{1,p}}, \ldots, r_{a(X_n)_{1,p}}]'$. There are as many unique **G** matrices as the number of unique multivariate contexts defined by the environmental covariates in **X**, yet the prediction of these matrices only requires estimating a much smaller set of CRN parameters. The log and inverse hyperbolic tangent link functions are respectively used to infer these trait-specific parameters (additive fixed effects, including global intercepts) defined on the transformed linear scale of genetic variances $\left[\boldsymbol{\beta}^\top_{\sigma^2_1}, \ldots, \boldsymbol{\beta}^\top_{\sigma^2_p}\right]^\top$ and genetic correlations $\left[\boldsymbol{\beta}^\top_{r_1}, \ldots, \boldsymbol{\beta}^\top_{r_{p-1,p}}\right]^\top$. In the general case, there will be $\operatorname{ncol}(\boldsymbol{X}) * p$ CRN parameters for genetic variances and $\operatorname{ncol}(\boldsymbol{X}) * \frac{p(p-1)}{2}$ parameters for the genetic correlations. Note that $\operatorname{atanh}(r) = \operatorname{logit}\left(\frac{r+1}{2}\right)/2$ extends the logit transformation defined for probability scale values to the scale of correlation coefficients.

It is theoretically important to recognize that any non-zero fixed effects predicting $\boldsymbol{G}_{(X)}$ provide evidence for gene-by-environment (GxE) interaction. In general, however, direct interpretation of these CRN fixed effect sizes will be challenging due to the distinct scales of link functions used for genetic variances and correlations. Therefore, once the model is estimated, I encourage researchers to use model predictions from Eq. 2 for more directly visualizing and quantifying total environmental effects on the more intuitive scales of genetic variances, correlations, and covariances, where $\sigma_{a(X_n)_{1,p}} = r_{a(X_n)_{1,p}}\sigma_{a(X_n)_1}\sigma_{a(X_n)_p}$. A worked example is provided below. When relevant, the same approach outlined above can also be taken to predict continuous and/or discrete effects on residual or permanent environmental (co)variances.

Extensions

It is important to emphasize that the basic CRN model presented here has been simplified to aid comprehension but can be flexibly extended using all of the standard tools available in a multilevel regression model. For example, further link functions can be used to generalize the linear models

presented above and predict the phenotypic means of non-Gaussian responses (using e.g. binomial, Poisson, gamma, beta, etc. distributions). Random individual-level slopes can also be introduced to the model for investigating the relationship between CRNs and RNs of phenotypic means, such as to understand how the genetic integration between phenotypic intercepts and slopes changes across developmental or social environments (Kraft et al., 2006; Stamps et al., 2018; Dingemanse et al., 2020; Bucklaew & Dochtermann, 2021; Martin et al., 2023). Nonlinear effects on the link transformed scales of genetic (co)variances can be predicted using standard polynomial approaches or using more flexible non-parametric methods and generalized additive functions such as splines or Gaussian processes (Pedersen et al., 2019; Riutort-Mayol et al., 2022), such as to incorporate spatiotemporal autocorrelation. It also possible to predict CRNs using random rather than fixed effects when environmental states are not directly measured (e.g. to capture random among-year variation). Such extensions are within the scope of any regression analysis and thus do not require specific attention with respect to the CRN model.

Phenotypic analysis

Many studies may lack the genetic information necessary to estimate **Eq. 2** or otherwise be principally interested in estimating phenotypic (co)variances. Without genetic data or repeated measurements, among- and within-individual patterns of phenotypic (co)variance will be confounded together, potentially biasing evolutionary predictions with measurement error and ephemeral environmental effects (Dingemanse et al., 2021; Martin, 2021). However, if multiple measurements are made on the same subjects across time, then repeatable among-individual differences in phenotype, due to both genetic variation and permanent environmental effects, can be effectively partitioned from stochastic variation using individual-level random effects. Following the same notation used above, but now introducing an *n* x *i* matrix $\boldsymbol{Y}$ to index repeated individual-level measurements, the phenotypic CRN is given by

$$\begin{bmatrix} \boldsymbol{z_1} \\ \vdots \\ \boldsymbol{z_p} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_1 + \boldsymbol{Y\mu}_{(X)_1} + \boldsymbol{\epsilon_1} \\ \vdots \\ \boldsymbol{X\beta}_p + \boldsymbol{Y\mu}_{(X)_p} + \boldsymbol{\epsilon_p} \end{bmatrix} \tag{3}$$

$$\begin{bmatrix} \boldsymbol{\mu}_{(X)_1} \\ \vdots \\ \boldsymbol{\mu}_{(X)_p} \end{bmatrix} \sim \boldsymbol{N}\left(\boldsymbol{0}, \boldsymbol{P}_{(X)}\right);\ \boldsymbol{P}_{(X_n)}: \begin{bmatrix} \sigma^2_{(X_n)_1} & \cdots & r_{(X_n)_{1,p}}\sigma_{(X_n)_1}\sigma_{(X_n)_p} \\ & \ddots & \vdots \\ & & \sigma^2_{(X_n)_p} \end{bmatrix}$$

$$\begin{bmatrix} \log\left(\boldsymbol{\sigma^2_{(X)_1}}\right) \\ \vdots \\ \log\left(\boldsymbol{\sigma^2_{(X)_p}}\right) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_{\boldsymbol{\sigma^2_1}} \\ \vdots \\ \boldsymbol{X\beta}_{\boldsymbol{\sigma^2_p}} \end{bmatrix};\quad \begin{bmatrix} \operatorname{atanh}\left(\boldsymbol{r}_{(X)_{1,2}}\right) \\ \vdots \\ \operatorname{atanh}\left(\boldsymbol{r}_{(X)_{p-1,p}}\right) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_{\boldsymbol{r_1}} \\ \vdots \\ \boldsymbol{X\beta}_{\boldsymbol{r_{p-1,p}}} \end{bmatrix}$$

where $\left[\boldsymbol{\mu}^{\top}_{(X)_1}, \ldots, \boldsymbol{\mu}^{\top}_{(X)_P}\right]^{\top}$ are random effects that are assumed to be independently distributed among individuals. As above, $\boldsymbol{P}_{(X_n)}$ is the matrix of among-individual phenotypic (co)variances predicted in response to the environmental context of measurement *n* for subject *i*, as determined by the CRN fixed effect parameters for phenotypic variances $\left[\boldsymbol{\beta}^{\top}_{\boldsymbol{\sigma^2_1}}, \ldots, \boldsymbol{\beta}^{\top}_{\boldsymbol{\sigma^2_p}}\right]^{\top}$ and correlations $\left[\boldsymbol{\beta}^{\top}_{\boldsymbol{r_1}}, \ldots, \boldsymbol{\beta}^{\top}_{\boldsymbol{r_p}}\right]^{\top}$. As with the quantitative genetic model, it is important to recognize that any non-zero fixed effects predicting $\boldsymbol{P}_{(X)}$ provide evidence for phenotype-by-environment (PxE) interactions. This same approach can also be used to extend the quantitative genetic model to repeated measurements, as is demonstrated in the worked example below (see **data availability** for accompanying code).

Statistical implementation

Bayesian inference in Stan

The CRN model (**Eq. 2-3**) cannot currently be estimated using standard statistical software packages for multivariate animal models and multilevel models more generally, due to a lack of in-built functionality for expressing elements of covariance matrices as generalized linear predictors. Fortunately, however, the extremely flexible Stan statistical programming language can be used to construct bespoke animal models of desired complexity within a Bayesian inferential framework, facilitating general estimation of CRNs models using cutting-edge Markov Chain Monte Carlo (MCMC) methods (Hoffman & Gelman, 2011; Nishio & Arakawa, 2019; Martin & Jaeggi, 2022). Detailed discussion of contemporary Bayesian statistics is beyond the scope of this paper. However, I encourage readers to consult up-to-date primers on Bayesian data analysis (Gelman et al., 2013, 2020; McElreath, 2020) for thorough introductions, including extensive tips and suggestions for key decisions such as the choice of priors, model validation and comparison, variable selection, and the interpretation of posterior estimates. As a general rule of thumb, I suggest using weakly regularizing priors when estimating CRN models, to reduce the risk of inferential bias while promoting more efficient model convergence (Lemoine, 2019; McElreath, 2020). Finally, note that despite it still being common to see thinning of MCMC chains reported in the literature, this practice is unnecessary and computationally inefficient (Link & Eaton, 2011).

Computational efficiency

This section covers formal details on efficient implementation of CRN models in Stan, which can be safely overlooked by empiricists without impeding interpretation or practical implementation. Prediction of large covariance matrices is computationally burdensome in a Bayesian framework, even with the use of appropriately regularizing priors and efficient MCMC algorithms. This reflects the fact that the probability of observing a permissible (i.e. positive-definite) covariance or correlation matrix declines rapidly with increasing dimensionality of the matrix (Dean & Majumdar, 2008). Estimation of the CRN model with three or more traits can, therefore, be best achieved through use of a mathematically equivalent but more computationally efficient reparameterization of the $\boldsymbol{G}_{(\boldsymbol{X})}$ and $\boldsymbol{P}_{(\boldsymbol{X})}$ matrices than is described by the standard parameterization presented in **Eq. 2-3**.

Firstly, the *p* x *p* correlation matrix $\boldsymbol{R}_{\boldsymbol{a}}$ containing all genetic (or phenotypic) correlations for *p* phenotypes can be decomposed using a Cholesky factorization such that

$$\boldsymbol{R}_{\boldsymbol{a}} = \boldsymbol{L}_{\boldsymbol{R}}\boldsymbol{L}_{\boldsymbol{R}}^{\top} \tag{4}$$

where $\boldsymbol{L}_{\boldsymbol{R}}$ is a lower-triangular matrix with unit length rows and positive diagonal elements. These assumptions reduce the number of free parameters necessary for calculating $\boldsymbol{R}_{\boldsymbol{a}}$, as the diagonal elements of $\boldsymbol{L}_{\boldsymbol{R}}$ are determined by the off-diagonal elements of each row. Therefore, estimating $\boldsymbol{L}_{\boldsymbol{R}}$ and subsequently deriving $\boldsymbol{R}_{\boldsymbol{a}}$ using **Eq. 4** improves computational time of the model (Stan Development Team, 2023). Following previous work on the prediction of covariance matrices (Lewandowski et al., 2009; Bloome & Schrage, 2021), computational efficiency can then be further increased by decomposing $\boldsymbol{L}_{\boldsymbol{R}}$ into a vector $\boldsymbol{\omega}$ of length $\frac{p(p-1)}{2}$ containing the canonical partial correlations constitutive of all unique lower-triangular elements in this matrix. The canonical partial correlations in $\boldsymbol{\omega}$ are of the same sign as their corresponding elements in $\boldsymbol{L}_{\boldsymbol{R}}$, but their magnitudes represent residual correlations between corresponding row and column variables after regressing both on all prior occurring row variables. In the general case, the canonical partial correlation ω_u, where $u = \frac{2cp-c^2+2r-3c-2}{2}$ is the vector element corresponding to unique lower-triangular Cholesky factor $L_{R[r,c]}$ at row *r* and column *c*, is given by

$$\omega_u = \begin{cases} L_{R[r,c]}, & \text{if } c = 1 < r \\ L_{R[r,c]} \,/\, \left(1 - \sum L^2_{R[r,1:c-1]}\right)^{\frac{1}{2}}, & \text{if } 1 < c \leq r \end{cases} \quad (5.1)$$

such that the Cholesky factor can in turn be derived from ω_u by

$$L_{R[r,c]} = \begin{cases} \omega_u, & \text{if } c = 1 < r \\ \omega_u * \left(1 - \sum L^2_{R[r,1:c-1]}\right)^{\frac{1}{2}}, & \text{if } 1 < c \leq r \end{cases} \quad (5.2)$$

This general decomposition strategy can be adapted for the CRN model by extending each element in the vector $\boldsymbol{\omega}$ to its own vector of context-specific canonical partial correlations. Using the same strategy developed above (Eq. 2-3), continuous environmental effects can then be specified and estimated more efficiently as predictors of the transformed canonical partial correlations

$$\begin{bmatrix} \operatorname{atanh}\left(\boldsymbol{\omega}_{(\mathbf{X})\mathbf{1}}\right) \\ \vdots \\ \operatorname{atanh}\left(\boldsymbol{\omega}_{(\mathbf{X})\frac{p(p-1)}{2}}\right) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_{\boldsymbol{\omega}_1} \\ \vdots \\ \boldsymbol{X\beta}_{\boldsymbol{\omega}_{\frac{p(p-1)}{2}}} \end{bmatrix} \quad (6)$$

Applying the inverse link function $\tanh()$ and using Eq. 5.2 to calculate Cholesky factorized matrices $\boldsymbol{L}_{\boldsymbol{R}(\boldsymbol{X})}$, the original context-specific correlation matrices can then be derived $\boldsymbol{R}_{\boldsymbol{a}(\boldsymbol{X})}$ and subsequently applied to generate model predictions for estimating environmental effects on a more familiar scale. In the presence of covarying environmental predictors, the CRN fixed effects $\boldsymbol{\beta}_{\boldsymbol{\sigma}^2}$ and $\boldsymbol{\beta}_{\boldsymbol{\omega}}$ (or $\boldsymbol{\beta}_{\boldsymbol{r}}$) can also be more efficiently estimated using a so-called thin QR factorization of the **X** matrix (Harville, 1997). This involves decomposing the predictor matrix $\boldsymbol{X} = \boldsymbol{Q}^*\boldsymbol{R}^*$ into an orthogonal matrix $Q^* = Q\sqrt{n-1}$ and upper-triangle matrix $R^* = \frac{R}{\sqrt{n-1}}$, estimating trait-specific regression coefficients using the orthogonal vectors $\boldsymbol{Q}^*\boldsymbol{\beta}^*$, and then returning regression coefficients appropriately scaled to the original data scale of $\boldsymbol{X}$ using $\boldsymbol{\beta} = \boldsymbol{R}^{*-\mathbf{1}}\boldsymbol{\beta}^*$. The QR decomposition increases efficiency by reducing posterior correlations during model sampling caused by covariation among measured predictors.

Finally, the Cholesky matrices $\boldsymbol{L}_{\boldsymbol{R}(\boldsymbol{X})}$ can also be used to more efficiently predict individual's context-specific additive genetic values from the CRN model. Following previous work by Martin and Jaeggi (2022), this can be accomplished using a matrix normal sampling distribution (Dutilleul, 1999), which extends the vectorized multivariate normal distribution to the sampling of multivariate normally distributed matrices. Using a n x p matrix $\boldsymbol{Z}_{\boldsymbol{G}}$ of standardized individual-level additive genetic deviations (i.e. z-scores of breeding values), a lower-triangular Cholesky decomposition $\boldsymbol{L}_{\boldsymbol{A}}$ of the relatedness matrix, and a diagonal matrix $\boldsymbol{S}_{\boldsymbol{a}(\boldsymbol{X}_n)} = \operatorname{diag}\left(\left[\sigma_{a(X_n)_1}, \ldots, \sigma_{a(X_n)_p}\right]\right)$ of context-specific genetic standard deviations, an *n* x *p* matrix of context-specific genetic values for each phenotype can be predicted by

$$\left[\boldsymbol{a}_{(X_n)_1}, \ldots, \boldsymbol{a}_{(X_n)_p}\right] = \mathbf{L}_{\mathbf{A}}\boldsymbol{Z}_{\boldsymbol{G}}\left(\mathbf{S}_{\mathbf{a}(\mathbf{X}_\mathbf{n})}\boldsymbol{L}_{\boldsymbol{R}(\boldsymbol{X}_n)}\right)^{\top} \sim \text{Matrix Normal}\left(\mathbf{0}_{\text{nxp}}, \boldsymbol{A}, \boldsymbol{G}_{(\boldsymbol{X}_n)}\right) \quad (7)$$

$$\rightarrow \operatorname{vec}\left(\left[\boldsymbol{a}_{(X_n)_1}, \ldots, \boldsymbol{a}_{(X_n)_p}\right]\right) \sim \boldsymbol{N}\left(\mathbf{0}, \boldsymbol{G}_{(\boldsymbol{X}_n)} \otimes \boldsymbol{A}\right)$$

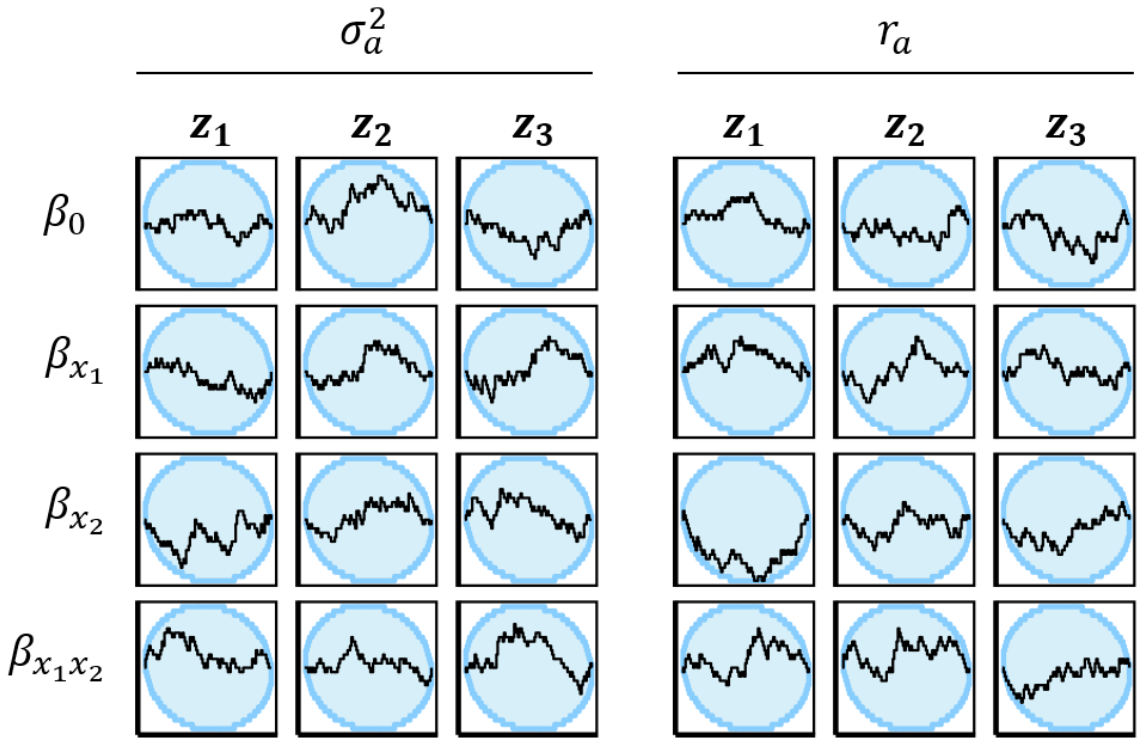
Easy-to-use R functions are provided (see **data availability**) to straightforwardly facilitate computational gains from Eq. 5-7 while also generating more intuitive model estimates and predictions with respect to the standard parameterization of the CRN model (Eq. 2-3).

Model validation

To provide a general validation of the proposed model, I conducted a simulation-based calibration (SBC) procedure to assess whether the quantitative genetic CRN (**Eq. 2**) is an unbiased Bayesian estimator. Note that the phenotypic CRN (**Eq. 3**) is simply a variant of the quantitative genetic model with independent random effects and thus does not require additional validation. SBC is a procedure for assessing the performance of a Bayesian algorithm across a broad range of possible parameter values generated from the prior distributions of a generative model (see Talts et al., 2018 for further details). This approach removes the need for arbitrarily picking a limited range of effect sizes for assessing performance and reduces the risk of missing unexpected sources of bias in uninvestigated regions of parameter space. Visual inspection of the correspondence between the generative prior distributions (expected values) and subsequent posterior distributions (inferred values) estimated during SBC is used to detect sources of bias, such as overdispersion in the estimator or inconsistent performance for extreme values.

100 datasets were simulated for SBC under very minimal sampling conditions of 100 individuals with a single measurement of 3 traits. Measurements were taken across environments characterized the interaction between 10 measured values of a continuous covariate (e.g. monthly temperatures, ages, plot densities) and a discrete factor with 2 levels (e.g. sex, dominance, breeding season yes/no). The interaction of these factors created 20 distinct environmental contexts; assuming balanced sampling, this resulted in 5 individuals being observed per context. Parameter values were generated using standard weakly regularizing priors (Lemoine, 2019; McElreath, 2020), such that $\boldsymbol{\beta} \sim N(0,1)$ for RN fixed effects determining phenotypic means and genetic (co)variances, $\boldsymbol{\sigma}_{\boldsymbol{\epsilon}} \sim \text{exponential}(2)$ for residual standard deviations, and $\boldsymbol{R}_{\boldsymbol{\epsilon}} \sim \text{LKJ}(2)$ for residual correlation matrices, which implied a zero-centered $\frac{r_{\alpha}+1}{2} \sim \text{Beta}(2,2)$ marginal distribution for each of the 3 correlations. Relatedness matrices were simply positive-definite correlation matrices simulated from $\boldsymbol{A} \sim \text{LKJ}(1)$. Posteriors for each dataset were estimated using 2000 MCMC samples across 4 chains using 500 samples each for warmup. Results from the SBC analysis showed that the distributions of inferred parameter values were congruent with the distributions of expected parameter values across the CRN fixed effects predicting genetic (co)variances (**Fig. 2**). In other words, posterior inferences were not systematically upwardly or downwardly biased from the true values used to generated the data. This provides strong evidence that the proposed Bayesian estimator provides unbiased inference of CRNs even under conditions of very minimal sampling effort and a reasonably broad range of effect sizes. It is important to emphasize that these results concern bias per se in estimates of expected values and do not quantify the statistical uncertainty or power of hypothesis tests for detecting these effects. Achieving high levels of power and low levels of uncertainty will generally require much larger sample sizes, as it is the case for any quantitative genetic analysis. Simulations functions are provided (see **data availability**) to aid researchers in carrying out a priori power analyses for effect and sample size ranges of interest.

Figure 2. Simulation-based calibration of the CRN model.



Footnote. Results are shown for analyses of 100 simulated datasets of 3 traits generated from prior distributions defined over the parameters of the quantitative genetic CRN model (Eq. 2). Plots show the difference between the expected cumulative density functions for CRN parameters of the genetic variances σ_a^2 and correlations r_a, based on their generative prior distributions, and the estimated cumulative density functions based on inferred posterior distributions of parameter values. The CRNs contained four parameters for genetic variances and correlations: β_0 (trait-specific intercepts), β_{x_1} (main effect of continuous environment), β_{x_2} (main effect of discrete environment), and $\beta_{x_1x_2}$ (interaction effect of continuous and discrete environments). Blue circles show 90% Bayesian credible intervals (0.95 probability) for regions of concordance between the estimated and expected parameter distributions, and the black line reflects the observed difference between the expected and inferred distribution (a perfectly horizontal line would thus indicate perfect concordance with the simulated parameters in every dataset). Consistent deviations of the black line beyond the blue region would provide evidence of systematic inferential bias during model estimation. Note that due to stochasticity, fluctuations of the black line within the blue circle are expected at computationally efficient sample sizes.

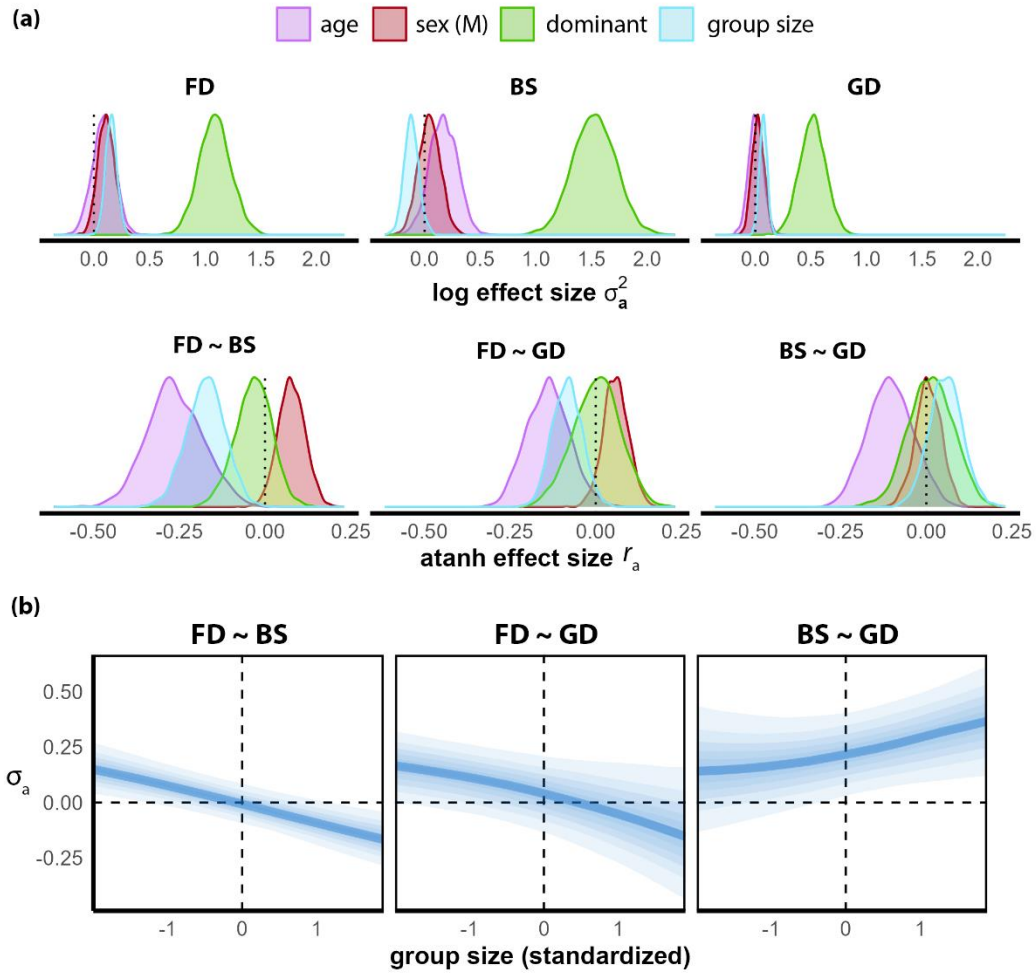
Worked example: cooperative behavior in meerkats

To demonstrate the utility of the proposed framework, I applied a CRN model to analyze an openly available dataset from a long-term study on the heritability of three cooperative behaviors (babysitting, pup feeding, and guarding/sentinel activity) in wild meerkats (Houslay et al., 2021). The goal of the analysis was to investigate how group size affects the genetic (co)variance of these cooperative behaviors, testing the hypothesis that task generalization decreases / specialization increases in larger groups (e.g. Bonner, 2004; Jeanson et al., 2007; Ulrich et al., 2018; Martin et al., 2023; **Figure 1d**). Using only data of individuals with measures available for all three behaviors, the total sample size for the analysis was 1560 pedigreed individuals with 6751 (babysitting), 6461 (pup feeding), and 11532 (guarding/sentinel activity) total observations. I simplified certain components of the animal models employed by these authors to focus attention on the CRN, using only the covariates (age, sex, dominance status, group size) that were available for all traits and were identified as important for understanding mean phenotypic differences in the meerkats' behavior. Additional random effects were included for each trait to capture individual-level permanent environmental effects, group identity during observation, breeding season, and individual x season interactions. The three phenotypes were modeled using binomial (half-days observed babysitting/total days) and Poisson (count of pup feeding and minutes in sentinel activity) distributions. Following **Eq. 2** and using the computational strategy explained in **Eq. 4-7**, the same environmental covariates used to predict phenotypic means were also used to predict potential changes in quantitative genetic (co)variances among these cooperative behaviors. Note that from the perspective of a gene, organismal attributes such as sex, age, and dominance (serving as proxies for various attendant changes in hormonal activity, social experiences, etc.) are just as much aspects of 'the environment' potentially modulating its expression as more exogenous factors like group size (e.g. Wittman et al., 2021). These covariates also allowed for appropriately testing the independent (age, sex, and dominance adjusted) effect of group size on genetic (co)variances among cooperative behaviors. A coding tutorial accompanying this worked example is provided on Github (see **data availability**).

The CRN analysis uncovered continuous changes in the genetic (co)variances of the meerkats' cooperative behaviors in response to group size, age, dominance, and sex, providing clear evidence for GxE interactions. These effects are visualized in **Fig. 3a** and summarized here using posterior medians and posterior probabilities of a directional effect p_+ / p_- (i.e. the probability of an effect being observed in the direction of the median effect size). As shown in **Fig. 3a**, the genetic variance of babysitting behavior (BS) was found to increase modestly with age (median $\beta_{\sigma^2} = 0.17, p_+ = 0.90$), while age had only weak and highly uncertain effects on pup feeding (FD: $\beta_{\sigma^2} = 0.09, p_+ = 0.81$) and vigilant guarding behavior (GD: $\beta_{\sigma^2} = 0.01, p_+ = 0.54$). Clearer evidence was provided for age reducing genetic correlations across behaviors (FD~BS: $\beta_r = -0.27, p_- = 0.99$; FD~GD: $\beta_r = -0.15, p_- = 0.99$; BS~GD: $\beta_r = -0.11, p_- = 0.94$). Males exhibited higher genetic variance in FD ($\beta_{\sigma^2} = 0.11, p_+ = 0.93$) but did not meaningfully differ from females for genetic variance in BS ($\beta_{\sigma^2} = 0.04, p_+ = 0.66$) or GD ($\beta_{\sigma^2} = 0.02, p_+ = 0.68$). Males also exhibited higher genetic correlations for FD~BS ($\beta_r = 0.07, p_+ = 0.96$) and FD~GD ($\beta_r = 0.06, p_+ = 0.95$) but did not differ in BS~GD ($\beta_r = 0.02, p_+ = 0.55$). Strong evidence was found for greater genetic variance among dominant individuals across behaviors (FD: $\beta_{\sigma^2} = 1.10, p_+ = 1.00$; BS: $\beta_{\sigma^2} = 1.54, p_+ = 1.00$; GD: $\beta_{\sigma^2} = 0.52, p_+ = 1.00$) in comparison to subordinates. However, dominance status had little to no effect on genetic correlations (FD~BS: $\beta_r = -0.03, p_- = 0.73$; FD~GD: $\beta_r = 0.01, p_+ = 0.55$; BS~GD: $\beta_r = 0.02, p_+ = 0.62$). Independently of these effects, group size was found to increase genetic variance in FD ($\beta_{\sigma^2} = 0.16, p_+ = 0.99$) and GD ($\beta_{\sigma^2} = 0.08, p_+ = 0.99$) while reducing genetic variance in BS ($\beta_{\sigma^2} = -0.12, p_- = 0.96$). Consistent with theoretical predictions, group size also acted to reduce genetic correlations among FD and BS ($\beta_r = -0.17, p_- = 0.99$) as well as FD and GD ($\beta_r =$

$-0.06, p_{-} = 0.92$), while increasing the genetic correlation among BS and GD ($\beta_r = 0.06, p_{+} = 0.90$). These results are combined together in **Fig. 3b** to show the genetic CRN of group size across behaviors. As can be seen, small group sizes tend to promote more generalized cooperative strategies, with positive genetic covariance across behaviors; larger group sizes in turn promote more specialized strategies, with negative covariance among genetic effects on FD and BS and GD, such that individuals who exhibit high levels of FD tended to exhibit lower levels of BS and GD (and vice versa).

Figure 3. Covariance reaction norms for meerkat social behavior.



Footnote. Posterior estimates are shown for the effect of environmental covariates on the among-individual genetic variances σ_a^2 of correlations r_a and (co)variances σ_a between meerkats' *pup feeding* (FD), *babysitting* (BS), and *guarding/sentinel* behavior (GD). (**a**) Scaled posterior densities are shown for the effects of age (units of months, standardized), sex, dominance status, and group size (units of 5, standardized) on genetic variances (log scale effects, top row) and correlations (atanh scale effects, bottom row). The vertical dotted line indicates zero effect, with values greater or less than zero indicating GxE interactions. (**b**) CRNs for the continuous, standardized effect of group size on genetic covariances, adjusted for the effects of sex, age, and dominance status. Posterior median effect sizes are shown by darker lines, with the surrounding shaded bands indicating 10 to 90% posterior credible intervals (CIs) from the darkest to lightest bands, respectively. Dotted vertical lines indicate the expected covariance at the average group size (0), while dotted horizontal lines indicate 0 genetic covariance, such that values above this line provide evidence for positive genetic covariance (generalization) and values below provide evidence for negative genetic covariance (specialization).

Conclusion

A longstanding goal unifying diverse fields of evolutionary science is to understand the role of developmental and contextual plasticity in the adaptation of complex phenotypes (Via et al., 1995; Paenke et al., 2007; Hutchings, 2011; Kuzawa & Bragg, 2012; Hendry, 2016; Pfennig, 2021). While strong theoretical emphasis has been placed on understanding the role of trait (co)variances in constraining multivariate evolution (Phillips & Arnold, 1989; Walsh & Blows, 2009; Chebib & Guillaume, 2017), it is often underappreciated that genetic and phenotypic (co)variances are themselves the product of underlying genotype- and phenotype-by-environment interactions (Service & Rose, 1985; Via & Lande, 1985; Pigliucci, 1996; Peiman & Robinson, 2017; Elgart et al., 2022; Martin et al., 2023). Modeling these dynamic environmental interactions is, therefore, a crucial but easily overlooked step in effectively explaining ongoing adaptation in a rapidly changing world (Westneat et al., 2019; Hudak & Dybdahl, 2023). Analytic tools for efficiently inferring these complex patterns have been limited, however, particularly outside of the laboratory, where organisms are exposed to continuous spatial and temporal variation in their local microhabitats. To the degree that such environmental variation is relevant for fitness and the benefits of responding to it outweigh the costs of producing a response, we expect for adaptive plasticity to evolve in trait expression (Gavrilets & Scheiner, 1993; de Jong, 1995; Haaland et al., 2021). In many cases, this plasticity will be reflected in average trait values; however, when fitness-relevant variation also occurs with respect to trait (co)variances within individuals' lifetimes (e.g. through fluctuating correlational selection, Revell, 2007; Roff & Fairbairn, 2012), adaptive plasticity may also evolve in trait variances and correlations (**Fig. 1**).

Important empirical efforts have been made to investigate the fluctuations in **G** and **P** matrices that result from such plasticity, as well as potentially rapid microevolution, in response to environmental heterogeneity and ongoing change in natural populations (e.g. Nespolo et al., 2009; Björklund et al., 2013; Bolund et al., 2015; Wood & Brodie, 2015). However, current character state approaches for analyzing changes in trait (co)variances rely on discretizing the environment. Outside of controlled experiments, this tends to encourage artificial binning of naturally occurring variation, which can reduce statistical power for true effects while also increasing the rate of false positives and downwardly biasing effect sizes (e.g. Cohen, 1983; MacCallum et al., 2002). Ultimately, this limits empiricists' ability to robustly infer continuous and potentially nonlinear environmental processes underlying such changes in the wild. The CRN model proposed here provides a novel and unbiased solution (**Fig. 2**) to this challenge, extending the standard animal model (Kruuk, 2004) to increase its flexibility in describing multivariate environmental effects on all aspects of genetic expression. As demonstrated by the worked example in meerkats, building on prior research by Houslay et al. (2021), CRNs can harness the rich information in long-term field datasets to generate fresh insights into longstanding theoretical questions, such as the role of group size in facilitating the emergence of niche specialization in animal societies (**Fig. 1, 3**). Further application of this novel approach is, therefore, likely to enhance our understanding of the evolution and ecology of multivariate plasticity across a variety of complex phenotypes in the wild.

Data availability

Guided tutorials for implementing CRNs, as well as R code for replicating the worked empirical example, are publicly available on Github at https://github.com/Jordan-Scott-Martin/covariance-reaction-norms .

References

Aguirre, J. D., Hine, E., McGuigan, K., & Blows, M. W. (2014). Comparing G: multivariate analysis of genetic variation in multiple populations. *Heredity*, *112*(1), 21–29.

Barneche, D. R., Robertson, D. R., White, C. R., & Marshall, D. J. (2018). Fish reproductive-energy output increases disproportionately with body size. *Science*, *360*(6389), 642–645.

Björklund, M., Husby, A., & Gustafsson, L. (2013). Rapid and unpredictable changes of the G-matrix in a natural bird population over 25 years. *Journal of Evolutionary Biology*, *26*(1), 1–13.

Björklund, Mats. (1996). The importance of evolutionary constraints in ecological time scales. *Evolutionary Ecology*, *10*(4), 423–431.

Björklund, Mats. (2004). Constancy of the G matrix in ecological time. *Evolution; International Journal of Organic Evolution*, *58*(6), 1157–1164.

Björklund, Mats, & Gustafsson, L. (2015). The stability of the G-matrix: The role of spatial heterogeneity. *Evolution; International Journal of Organic Evolution*, *69*(7), 1953–1958.

Bloome, D., & Schrage, D. (2021). Covariance Regression Models for Studying Treatment Effect Heterogeneity Across One or More Outcomes: Understanding How Treatments Shape Inequality. *Sociological Methods & Research*, *50*(3), 1034–1072.

Bolund, E., Hayward, A., Pettay, J. E., & Lummaa, V. (2015). Effects of the demographic transition on the genetic variances and covariances of human life-history traits. *Evolution; International Journal of Organic Evolution*, *69*(3), 747–755.

Bonner, J. T. (2004). Perspective: the size-complexity rule. *Evolution; International Journal of Organic Evolution*, *58*(9), 1883–1890.

Brommer, J., Class, B., & Covarrubias-Pazaran, G. (2019). Multivariate Mixed Models in Ecology and Evolutionary biology: Inferences and implementation in R. In *EcoEvoRxiv*. https://doi.org/10.32942/osf.io/hs38a

Brooks, R., Hunt, J., Blows, M. W., Smith, M. J., Bussière, L. F., & Jennions, M. D. (2005). Experimental evidence for multivariate stabilizing sexual selection. *Evolution; International Journal of Organic Evolution*, *59*(4), 871–880.

Bucklaew, A., & Dochtermann, N. A. (2021). The effects of exposure to predators on personality and plasticity. *Ethology: Formerly Zeitschrift Fur Tierpsychologie*, *127*(2), 158–165.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*. https://doi.org/10.18637/jss.v076.i01

Carvalho, M. J. A., & Mirth, C. K. (2015). Coordinating morphology with behavior during development: an integrative approach from a fly perspective. *Frontiers in Ecology and Evolution*, *3*. https://doi.org/10.3389/fevo.2015.00005

Chang, C.-C., Moiron, M., Sánchez-Tójar, A., Niemela, P., & Laskowski, K. (2023). What's the meta-analytic evidence for life-history trade-offs at the genetic level? In *Authorea*. https://doi.org/10.22541/au.168795190.00787983/v1

Chebib, J., & Guillaume, F. (2017). What affects the predictability of evolutionary constraints using a G-matrix? The relative effects of modular pleiotropy and mutational correlation. *Evolution; International Journal of Organic Evolution*, *71*(10), 2298–2312.

Cohen, J. (1983). The Cost of Dichotomization. *Applied Psychological Measurement*, *7*(3), 249–253.

Damián, X., Ochoa-López, S., Gaxiola, A., Fornoni, J., Domínguez, C. A., & Boege, K. (2020). Natural selection acting on integrated phenotypes: covariance among functional leaf traits increases plant fitness. *The New Phytologist*, *225*(1), 546–557.

Danchin, É., & Wagner, R. H. (2010). Inclusive heritability: combining genetic and non-genetic information to study animal behavior and culture. *Oikos* , *119*(2), 210–218.

de Jong, G. (1995). Phenotypic Plasticity as a Product of Selection in a Variable Environment. *The American Naturalist*, *145*(4), 493–512.

de la Mata, R., Zas, R., Bustingorri, G., Sampedro, L., Rust, M., Hernandez-Serrano, A., & Sala, A. (2022). Drivers of population differentiation in phenotypic plasticity in a temperate conifer: A 27-year study. *Evolutionary Applications*, *15*(11), 1945–1962.

Dean, D. S., & Majumdar, S. N. (2008). Extreme value statistics of eigenvalues of Gaussian random matrices. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *77*(4 Pt 1), 041108.

Dingemanse, Niels J., Barber, I., & Dochtermann, N. A. (2020). Non-consumptive effects of predation: does perceived risk strengthen the genetic integration of behaviour and morphology in stickleback? *Ecology Letters*, *23*(1), 107–118.

Dingemanse, Niels J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: mixed-effect modelling approaches. *The Journal of Animal Ecology*, *82*(1), 39–54.

Dingemanse, Niels Jeroen, Araya-Ajoy, Y. G., & Westneat, D. F. (2021). Most published selection gradients are underestimated: Why this is and how to fix it. *Evolution; International Journal of Organic Evolution*, *75*(4), 806–818.

Dutilleul, P. (1999). The mle algorithm for the matrix normal distribution. *Journal of Statistical Computation and Simulation*, *64*(2), 105–123.

Elgart, M., Goodman, M. O., Isasi, C., Chen, H., Morrison, A. C., de Vries, P. S., Xu, H., Manichaikul, A. W., Guo, X., Franceschini, N., Psaty, B. M., Rich, S. S., Rotter, J. I., Lloyd-Jones, D. M., Fornage, M., Correa, A., Heard-Costa, N. L., Vasan, R. S., Hernandez, R., … Sofer, T. (2022). Correlations between complex human phenotypes vary by genetic background, gender, and environment. *Cell Reports. Medicine*, *3*(12), 100844.

Estes, S., & Arnold, S. J. (2007). Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *The American Naturalist*, *169*(2), 227–244.

Ferguson-Gow, H., Sumner, S., Bourke, A. F. G., & Jones, K. E. (2014). Colony size predicts division of labour in attine ants. *Proceedings. Biological Sciences / The Royal Society*, *281*(1793). https://doi.org/10.1098/rspb.2014.1411

Fogarty, L., & Wade, M. J. (2022). Niche construction in quantitative traits: heritability and response to selection. *Proceedings. Biological Sciences / The Royal Society*, *289*(1976), 20220401.

Gavrilets, S., & Scheiner, S. M. (1993). The genetics of phenotypic plasticity. V. Evolution of reaction norm shape. *Journal of Evolutionary Biology*, *6*(1), 31–48.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis Third edition*. CRC Press.

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., & Modrák, M. (2020). Bayesian Workflow. In *arXiv [stat.ME]*. arXiv. http://arxiv.org/abs/2011.01808

Grebe, N. M., Fitzpatrick, C., Sharrock, K., Starling, A., & Drea, C. M. (2019). Organizational and activational androgens, lemur social play, and the ontogeny of female dominance. *Hormones and Behavior*, *115*, 104554.

Haaland, T. R., Wright, J., & Ratikainen, I. I. (2021). Individual reversible plasticity as a genotype-level bet-hedging strategy. *Journal of Evolutionary Biology*, *34*(7), 1022–1033.

Haave-Audet, E., Besson, A. A., Nakagawa, S., & Mathot, K. J. (2022). Differences in resource acquisition, not allocation, mediate the relationship between behaviour and fitness: a systematic review and meta-analysis. *Biological Reviews of the Cambridge Philosophical Society*, *97*(2), 708–731.

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. Springer-Verlag.

Heino, M., Díaz Pauli, B., & Dieckmann, U. (2015). Fisheries-Induced Evolution. *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 461–480.

Hendry, A. P. (2016). Key Questions on the Role of Phenotypic Plasticity in Eco-Evolutionary Dynamics. *The Journal of Heredity*, *107*(1), 25–41.

Henry, G. A., & Stinchcombe, J. R. (2023). G-matrix stability in clinally diverging populations of an annual weed. *Evolution; International Journal of Organic Evolution*, *77*(1), 49–62.

Hoffman, M. D., & Gelman, A. (2011). The no-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. arXiv. https://jmlr.org/papers/volume15/hoffman14a/hoffman14a.pdf

Houslay, T. M., Nielsen, J. F., & Clutton-Brock, T. H. (2021). Contributions of genetic and nongenetic sources to variation in cooperative behavior in a cooperative mammal. *Evolution; International Journal of Organic Evolution*, *75*(12), 3071–3086.

Huber, S. E., Lenz, B., Kornhuber, J., & Müller, C. P. (2017). Prenatal androgen-receptor activity has organizational morphological effects in mice. *PloS One*, *12*(11), e0188752.

Hudak, A., & Dybdahl, M. (2023). Phenotypic plasticity under the effects of multiple environmental variables. *Evolution; International Journal of Organic Evolution*, *77*(6), 1370–1381.

Hutchings, J. A. (2011). Old wine in new bottles: reaction norms in salmonid fishes. *Heredity*, *106*(3), 421–437.

Jeanson, R., Fewell, J. H., Gorelick, R., & Bertram, S. M. (2007). Emergence of Increased Division of Labor as a Function of Group Size. *Behavioral Ecology and Sociobiology*, *62*(2), 289–298.

Jusufovski, D., & Kuparinen, A. (2020). Exploring individual and population eco-evolutionary feedbacks under the coupled effects of fishing and predation. *Fisheries Research*, *231*, 105713.

Karsai, I., & Wenzel, J. W. (1998). Productivity, individual-level and colony-level flexibility, and organization of work as consequences of colony size. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(15), 8665–8669.

Koch, E. L., Sbilordo, S. H., & Guillaume, F. (2020). Genetic variance in fitness and its cross-sex covariance predict adaptation during experimental evolution. *Evolution; International Journal of Organic Evolution*, *74*(12), 2725–2740.

Kraft, P. G., Wilson, R. S., Franklin, C. E., & Blows, M. W. (2006). Substantial changes in the genetic basis of tadpole morphology of Rana lessonae in the presence of predators. *Journal of Evolutionary Biology*, *19*(6), 1813–1818.

Kruuk, L. E. B., & Hadfield, J. D. (2007). How to separate genetic and environmental causes of similarity between relatives. *Journal of Evolutionary Biology*, *20*(5), 1890–1903.

Kruuk, Loeske E. B. (2004). Estimating genetic parameters in natural populations using the 'animal model.' *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1446), 873–890.

Kuzawa, C. W., & Bragg, J. M. (2012). Plasticity in Human Life History Strategy: Implications for Contemporary Human Variation and the Evolution of Genus Homo. *Current Anthropology*, *53*(S6), S369–S382.

Lande, R. (1976). NATURAL SELECTION AND RANDOM GENETIC DRIFT IN PHENOTYPIC EVOLUTION. *Evolution; International Journal of Organic Evolution*, *30*(2), 314–334.

Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain body size allometry. *Evolution; International Journal of Organic Evolution*, *33*(1Part2), 402–416.

Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution; International Journal of Organic Evolution*, *37*(6), 1210–1226.

Lemoine, N. P. (2019). Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos* , *128*(7), 912–928.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001.

Link, W. A., & Eaton, M. J. (2011). On thinning of chains in MCMC. *Methods in Ecology and Evolution*, *3*(1). https://doi.org/10.1111/j.2041-210X.2011.00131.x

Lofeu, L., Brandt, R., & Kohlsdorf, T. (2017). Phenotypic integration mediated by hormones: associations among digit ratios, body size and testosterone during tadpole development. *BMC Evolutionary Biology*, *17*(1), 175.

Lynch, M., & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19–40.

Martin, J. (2021). Estimating nonlinear selection on behavioral reaction norms. In *EcoEvoRxiv*. https://doi.org/10.32942/osf.io/u26tz

Martin, J. S., & Jaeggi, A. V. (2022). Social animal models for quantifying plasticity, assortment, and selection on interacting phenotypes. *Journal of Evolutionary Biology*, *35*(4), 520–538.

Martin, J. S., Jaeggi, A. V., & Koski, S. E. (2023). The social evolution of individual differences: Future directions for a comparative science of personality in social behavior. *Neuroscience and Biobehavioral Reviews*, *144*, 104980.

McElreath, R. (2020). *Statistical Rethinking; A Bayesian Course with Examples in R and Stan; Second Edition*. CRC Press.

McGlothlin, J. W., Kobiela, M. E., Wright, H. V., Mahler, D. L., Kolbe, J. J., Losos, J. B., & Brodie, E. D., 3rd. (2018). Adaptive radiation along a deeply conserved genetic line of least resistance in Anolis lizards. *Evolution Letters*, *2*(4), 310–322.

Mitchell, D. J., & Houslay, T. M. (2021). Context-dependent trait covariances: how plasticity shapes behavioral syndromes. *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology*, *32*(1), 25–29.

Montoya, E. R., Terburg, D., Bos, P. A., Will, G.-J., Buskens, V., Raub, W., & van Honk, J. (2013). Testosterone administration modulates moral judgments depending on second-to-fourth digit ratio. *Psychoneuroendocrinology*, *38*(8), 1362–1369.

Nishio, M., & Arakawa, A. (2019). Performance of Hamiltonian Monte Carlo and No-U-Turn Sampler for estimating genetic parameters and breeding values. *Genetics, Selection, Evolution: GSE*, *51*(1), 73.

Nussey, D. H., Wilson, A. J., & Brommer, J. E. (2007). The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology*, *20*(3), 831–844.

Oh, K. P., & Shaw, K. L. (2013). Multivariate sexual selection in a rapidly evolving speciation phenotype. *Proceedings. Biological Sciences / The Royal Society*, *280*(1761), 20130482.

Oomen, R. A., & Hutchings, J. A. (2022). Genomic reaction norms inform predictions of plastic and adaptive responses to climate change. *The Journal of Animal Ecology*, *91*(6), 1073–1087.

Paenke, I., Sendhoff, B., & Kawecki, T. J. (2007). Influence of plasticity and learning on evolution under directional selection. *The American Naturalist*, *170*(2), E47-58.

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ*, *7*, e6876.

Pfennig, D. W. (2021). *Phenotypic Plasticity & Evolution* (1st Edition). CRC Press.

Phillips, P. C., & Arnold, S. J. (1989). Visualizing Multivariate Selection. *Evolution; International Journal of Organic Evolution*, *43*(6), 1209–1222.

Pigliucci, M. (1996). Modelling phenotypic plasticity. II. Do genetic correlations matter? *Heredity*, *77 (Pt 5)*, 453–460.

R Core Team. (2023). *R: A language and environment for statistical computing*.

Revell, L. J. (2007). The G matrix under fluctuating correlational mutation and selection. *Evolution; International Journal of Organic Evolution*, *61*(8), 1857–1872.

Riutort-Mayol, G., Bürkner, P.-C., Andersen, M. R., Solin, A., & Vehtari, A. (2022). Practical Hilbert space approximate Bayesian Gaussian processes for probabilistic programming. *Statistics and Computing*, *33*(1), 17.

Roff, D. A. (1996). The evolution of genetic correlations: An analysis of patterns. *Evolution; International Journal of Organic Evolution*, *50*(4), 1392–1403.

Roff, D. A., & Fairbairn, D. J. (2012). A test of the hypothesis that correlational selection generates genetic correlations. *Evolution; International Journal of Organic Evolution*, *66*(9), 2953–2960.

Rolian, C. (2020). Ecomorphological specialization leads to loss of evolvability in primate limbs. *Evolution; International Journal of Organic Evolution*, *74*(4), 702–715.

Royauté, R., Hedrick, A., & Dochtermann, N. A. (2020). Behavioural syndromes shape evolutionary trajectories via conserved genetic architecture. *Proceedings. Biological Sciences / The Royal Society*, *287*(1927), 20200183.

Schaum, C.-E., Buckling, A., Smirnoff, N., & Yvon-Durocher, G. (2022). Evolution of thermal tolerance and phenotypic plasticity under rapid and slow temperature fluctuations. *Proceedings. Biological Sciences / The Royal Society*, *289*(1980), 20220834.

Schluter, D. (1996). Adaptive Radiation Along Genetic Lines of Least Resistance. *Evolution; International Journal of Organic Evolution*, *50*(5), 1766–1774.

Service, Philip M., & Rose, M. R. (1985). Genetic Covariation Among Life-History Components: The Effect of Novel Environments. *Evolution; International Journal of Organic Evolution*, *39*(4), 943–945.

Sharpe, D. M. T., & Hendry, A. P. (2009). Life history change in commercially exploited fish stocks: an analysis of trends across studies. *Evolutionary Applications*, *2*(3), 260–275.

Stamps, J. A., Biro, P. A., Mitchell, D. J., & Saltz, J. B. (2018). Bayesian updating during development predicts genotypic differences in plasticity. *Evolution; International Journal of Organic Evolution*, *72*(10), 2167–2180.

Stan Development Team. (2023). *Stan Modeling Language Users Guide and Reference Manual 2.33*.

Stearns, S. C. (1989). Trade-Offs in Life-History Evolution. *Functional Ecology*, *3*(3), 259–268.

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., & Gelman, A. (2018). Validating Bayesian Inference Algorithms with Simulation-Based Calibration. In *arXiv [stat.ME]*. arXiv. http://arxiv.org/abs/1804.06788

Thomas, M. L., & Elgar, M. A. (2003). Colony size affects division of labour in the ponerine ant Rhytidoponera metallica. *Die Naturwissenschaften*, *90*(2), 88–92.

Thomson, C. E., Winney, I. S., Salles, O. C., & Pujol, B. (2018). A guide to using a multiple-matrix animal model to disentangle genetic and nongenetic causes of phenotypic variance. *PloS One*, *13*(10), e0197720.

Ulrich, Y., Saragosti, J., Tokita, C. K., Tarnita, C. E., & Kronauer, D. J. C. (2018). Fitness benefits and emergent division of labour at the onset of group living. *Nature*, *560*(7720), 635–638.

Uusi-Heikkilä, S. (2020). Implications of size-selective fisheries on sexual selection. *Evolutionary Applications*, *13*(6), 1487–1500.

van Noordwijk, A. J., & de Jong, G. (1986). Acquisition and Allocation of Resources: Their Influence on Variation in Life History Tactics. *The American Naturalist*, *128*(1), 137–142.

Via, S., Gomulkiewicz, R., De Jong, G., Scheiner, S. M., Schlichting, C. D., & Van Tienderen, P. H. (1995). Adaptive phenotypic plasticity: consensus and controversy. *Trends in Ecology & Evolution*, *10*(5), 212–217.

Via, Sara, & Lande, R. (1985). Genotype-Environment Interactions and the Evolution of Phenotypic Plasticity. *Evolution; International Journal of Organic Evolution*, *39*(3), 505–522.

vom Saal, F. S. (1979). Prenatal exposure to androgen influences morphology and aggressive behavior of male and female mice. *Hormones and Behavior*, *12*(1), 1–11.

Walsh, B., & Blows, M. W. (2009). Abundant Genetic Variation + Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 41–59.

Westneat, D. F., Potts, L. J., Sasser, K. L., & Shaffer, J. D. (2019). Causes and Consequences of Phenotypic Plasticity in Complex Environments. *Trends in Ecology & Evolution*, *34*(6), 555–568.

Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., Kruuk, L. E. B., & Nussey, D. H. (2010). An ecologist's guide to the animal model. *The Journal of Animal Ecology*, *79*(1), 13–26.

Wittman, T. N., Robinson, C. D., McGlothlin, J. W., & Cox, R. M. (2021). Hormonal pleiotropy structures genetic covariance. *Evolution Letters*, *5*(4), 397–407.

Wood, C. W., & Brodie, E. D., 3rd. (2015). Environmental effects on the structure of the G-matrix. *Evolution; International Journal of Organic Evolution*, *69*(11), 2927–2940.

Yewers, M. S. C., Jessop, T. S., & Stuart-Fox, D. (2017). Endocrine differences among colour morphs in a lizard with alternative behavioural strategies. *Hormones and Behavior*, *93*, 118–127.