

A Bayesian framework for comparative quantitative genetics

Otso Ovaskainen*, José Manuel Cano and Juha Merilä

Department of Biological and Environmental Sciences, P. O. Box 65,
F1-00014, University of Helsinki, Finland

Bayesian approaches have been extensively used in animal breeding sciences, but similar approaches in the context of evolutionary quantitative genetics have been rare. We compared the performance of Bayesian and frequentist approaches in estimation of quantitative genetic parameters (*viz.* matrices of additive and dominance variances) in datasets typical of evolutionary studies and traits differing in their genetic architecture. Our results illustrate that it is difficult to disentangle the relative roles of different genetic components from small datasets, and that ignoring, e.g. dominance is likely to lead to biased estimates of additive variance. We suggest that a natural summary statistic for **G**-matrix comparisons can be obtained by examining how different the underlying multinormal probability distributions are, and illustrate our approach with data on the common frog (*Rana temporaria*). Furthermore, we derive a simple Monte Carlo method for computation of fraternity coefficients needed for the estimation of dominance variance, and use the pedigree of a natural Siberian jay (*Perisoreus infaustus*) population to illustrate that the commonly used approximate values can be substantially biased.

Keywords: Bayesian estimation; comparative quantitative genetics; fraternity coefficient; pedigree; matrix comparison; probability metric

1. INTRODUCTION

The matrix of additive genetic variances and covariances (**G**) is fundamental in the study of evolutionary processes of natural selection and genetic drift shaping patterns of phenotypic (co)variation (Lande 1979; McGuigan 2006). Under the assumption that **G** is stable over time, it can be used to predict and reconstruct evolutionary trajectories if selection gradients are known (Lande & Arnold 1983; Schluter 1996; Arnold *et al.* 2001). While the link between the **G** matrix and evolutionary processes is well understood, difficulties in estimating **G** for natural populations have hindered the full usage of its potential in evolutionary research. Attempts to overcome the problem have led to heterogeneous methodologies in the estimation and comparison of **G** matrices (reviewed in Hofer (1998) and Steppan *et al.* (2002)). The diversity of the approaches (e.g. analysis of variance, minimum norm and minimum variance quadratic unbiased estimators, restricted maximum likelihood, principal component analyses, dimensionality analyses) has resulted in an inefficient use of data, lack of comparability across studies and difficulties in the biological interpretation (Steppan *et al.* 2002; Mezey & Houle 2003; Merilä & Björklund 2004).

In the context of evolutionary studies of natural populations, genetic variances were initially estimated by least-squares methods (Falconer & Mackay 1996; Roff 1997; Lynch & Walsh 1998). Subsequently, likelihood-based methods, such as maximum likelihood (ML),

restricted maximum likelihood (REML) and Bayesian methods have become preferred, since they can handle arbitrary breeding designs and can account for selection processes or make them ignorable (Shaw 1987; Blasco 2001). Among these, there is an increasing consensus towards the benefits of Bayesian modelling (Shoemaker *et al.* 1999; Sorensen & Gianola 2002; Beaumont & Rannala 2004), the major advantage being the ability to quantify uncertainty in multivariate problems by assessing the full joint posterior distribution of the model parameters (Gelman *et al.* 2004). This advantage becomes pronounced if the estimated **G** matrix is to be used in further analyses, such as assessing heritability or the response to selection. Once a posterior for **G** is available, the distribution of any variable that is a function of **G** can be calculated. Bayesian methods also avoid the need for bootstrapping, which is difficult especially in case of arbitrary pedigrees. The choice of a natural resampling unit can be problematic even in case of simple breeding designs (O'Hara & Merilä 2005).

A further advantage of the Bayesian approach relates to confounding variables. If the model includes two explanatory variables that are hard to disentangle from each other, ML methods can be very sensitive to small perturbations in the data. For example, in a study of the rainbow trout (Pante *et al.* 2002), in one population the effect of dominance was estimated to be zero, and the effect of environmental influences was non-zero. In another population, the effect of dominance was estimated to be non-zero, and the effect of environmental influences was zero. The effect of dominance is often difficult to tease apart from common environmental influences, as full-sibs typically experience the same environment. A Bayesian approach including both effects would give an accurate estimate for the sum of the two variance components, but include wide

* Author for correspondence (otso.ovaskainen@helsinki.fi).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2007.0949> or via <http://journals.royalsociety.org>. One contribution of 18 to a Special Issue 'Evolutionary dynamics of wild populations'.

uncertainty in the partitioning of the sum into the two components. The posterior distribution would hence correctly imply that there is variance at the family level, but that there is not much evidence in the data to tell whether the effect comes from dominance or from common environmental effects. Hence, while any statistical method cannot disentangle confounded variance components, a Bayesian approach describes the information content in the data in a robust and intuitive manner.

In spite of the advantages of Bayesian approaches, the usage of Bayesian models for estimation of genetic components is still scarce in the evolutionary context and has been mainly restricted to univariate cases (Pakkasmaa *et al.* 2003; Palo *et al.* 2003; Cano *et al.* 2004; Waldmann 2004). This is in striking contrast with the field of animal breeding, where Bayesian estimation schemes had been developed for multivariate models already more than two decades ago (Gianola & Fernando 1986) and are currently routinely used. The field is still actively developing, much of the current emphasis being in extensions of the multinormal modelling framework, including, e.g. binary and categorial response variables, thick-tailed models, right-ordered Gaussian traits and arbitrary patterns of missing data (Sorensen & Gianola 2002; Korsgaard *et al.* 2003). As animal breeders are mainly interested in the estimation of breeding values and genetic merit, originally little emphasis was paid to causal components other than additive effects. However, ignoring, e.g. maternal or dominance effects can lead to biased estimates of breeding values (Miształ *et al.* 1997), and the models used currently in animal breeding allow for the joint estimation of an arbitrary number of causal components. Given the maturity of the methodology developed in the context of animal breeding, the limited use of the Bayesian framework in the evolutionary context is somewhat surprising. With natural populations, there is usually much less data than in animal breeding, hence assessing uncertainty is of especial importance.

In this paper, we illustrate the use of the Bayesian framework by applying it to contrasting study designs common in evolutionary research, using both simulated and real data. With regard to simulated data, we consider a North Carolina II (NC II)-type breeding design and a long-term pedigree structure corresponding to a natural population of Siberian jay. With regard to real data, we consider NC II breeding data on the common frog. We develop a new approach to \mathbf{G} -matrix comparison by making the point that \mathbf{G} represents a multinormal distribution, and by using probability metrics to examine how different the underlying multinormal distributions are between two populations. We illustrate our approach with the common frog data, and compare with earlier results based on common principal component analysis (Cano *et al.* 2004). We start by noting that the estimation of dominance variance requires the calculation of fraternity coefficients, which are typically approximated by deriving them from the coancestry coefficients. We illustrate how the fraternity coefficients can be easily calculated (up to a given accuracy) using a straightforward simulation approach, and show that the true values can deviate substantially from the approximate ones in case of a long-term pedigree.

2. MATERIAL AND METHODS

We apply the Bayesian framework to contrasting study designs common in evolutionary research (figure 1). The first (figure 1a) is a pedigree of a natural population of the Siberian jay (*Perisoreus infaustus*), totalling 18 generations and 1092 individuals, with on average at least one great grandparent known for 4.6 generations in the past (Lillandt *et al.* 2003). As common with natural populations, every generation includes a number of immigrants from outside the study area and hence of unknown origin. The second (figure 1b) is a NC II-type breeding design (Kearsey & Pooni 1996), consisting of blocks of sires crossed with dams.

(a) Calculation of coancestry and fraternity coefficients

To estimate the additive and dominance components, coefficients of coancestry and fraternity are required. While fraternity coefficients can be calculated directly for designs such as NC II, simple direct methods lead to approximations in case of complex pedigrees. We illustrate how the correct values of the fraternity coefficients can be easily calculated up to desired accuracy by a straightforward simulation approach, and examine in the case of the Siberian jay pedigree if the true values deviate substantially from the ones obtained by a commonly used direct method.

We denote by θ_{ij} the coefficient of coancestry, defined as the probability that randomly chosen genes from individuals i and j are identical by descent. The additive genetic relationship matrix \mathbf{A} is then defined with elements $A_{ij} = 2\theta_{ij}$. The matrix can be calculated recursively using the formulae, $A_{ii} = 1 + A_{m_i, f_i}/2$, $A_{ij} = (A_{j, m_i} + A_{j, f_i})/2$, where m_i and f_i refer to mother and father of individual i , respectively (Lynch & Walsh 1998). The coefficient of fraternity Δ_{ij} is defined as the probability that both alleles of individuals i and j are identical by descent. This probability is commonly derived from the coefficients of coancestry by the formula (Lynch & Walsh 1998)

$$\Delta_{ij} \approx \theta_{m_i, m_j} \theta_{f_i, f_j} + \theta_{m_i, f_j} \theta_{f_i, m_j}. \quad (2.1)$$

Equation (2.1) is based on the logic that there are two alternatives by which both genes can be identical by descent. Let us denote by \mathbf{X} the event represented by the first term, which is the case if the gene descending from m_i is identical by descent with that descending from m_j (event \mathbf{X}_1), and if the gene from f_i is identical by descent with that from f_j (event \mathbf{X}_2). The second term represents event \mathbf{Y} , in which the genes from m_i and f_j are identical by descent (event \mathbf{Y}_1), and so are the genes from f_i and m_j (event \mathbf{Y}_2). However, equation (2.1) is not exact, as it assumes that the four probabilities are independent of each other. To see this, we note first that $\Delta_{ij} = \Pr(\mathbf{X} \text{ or } \mathbf{Y})$, which can be written as $\Delta_{ij} = \Pr(\mathbf{X}) + \Pr(\mathbf{Y}) - \Pr(\mathbf{X} \text{ and } \mathbf{Y})$, out of which equation (2.1) ignores the intersection. Second, $\Pr(\mathbf{X}) = \Pr(\mathbf{X}_1 \text{ and } \mathbf{X}_2) = \Pr(\mathbf{X}_1)\Pr(\mathbf{X}_2|\mathbf{X}_1)$ (a similar identity holds for $\Pr(\mathbf{Y})$), out of which equation (2.1) ignores the conditional part. As direct recursive methods for the calculation of the exact fraternity values can be very slow for long pedigrees, typically compromises have been made, e.g. using equation (2.1), ignoring dominance connections through great grandparents (Miształ 1997), or including the inbreeding coefficient in the animal model (de Boer & van Arendonk 1992). However, such compromising can be avoided by computing the fraternity coefficients up to a required accuracy using a straightforward simulation

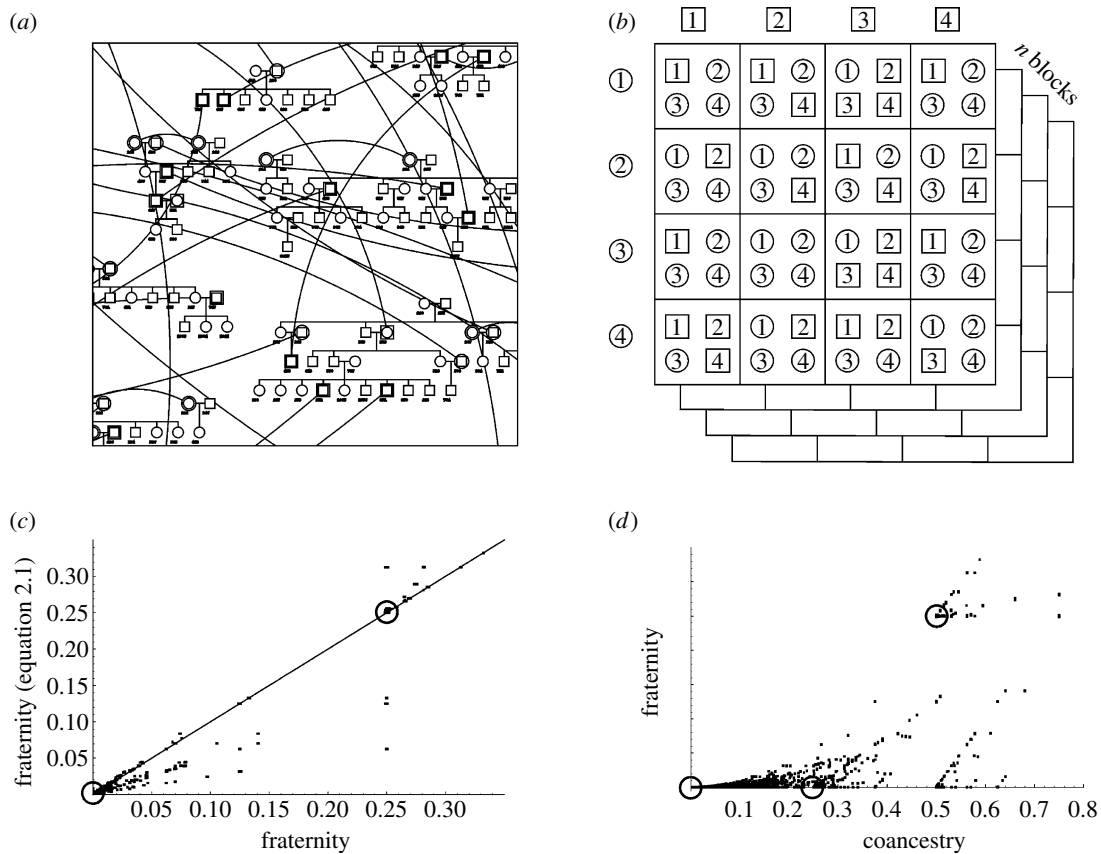


Figure 1. Data structures typical of evolutionary studies. (a) Part of the pedigree for a natural Siberian jay population and (b) a North Carolina type II breeding design. In (a,b), squares depict males and circles depict females. (c) Comparison of the true values of the fraternity coefficients (computed for the Siberian jay pedigree by the method of appendix A in the electronic supplementary material, standard error at most s.e.=0.0005) with approximations obtained by equation (2.1). (d) Plots of fraternity coefficients against coancestry coefficients. In (c,d), the dots correspond to the Siberian jay pedigree and the circles to the NC II breeding design. Self-pairs (A_{ii} and Δ_{ii}) are omitted in (c,d).

approach (appendix A in the electronic supplementary material).

(b) Estimation of genetic components

We consider the standard multivariate linear additive genetic model, which allows for an arbitrary number of causal components (see appendix B in the electronic supplementary material), but which we restrict in this section for notational simplicity to additive and dominance effects only. Let \mathbf{y} denote the data available on k traits of n animals. Denoting by $y_{i,j}$ the trait j of individual i , we order \mathbf{y} as $\mathbf{y} = (y_{1,1}, y_{2,1}, \dots, y_{n,1}, y_{1,2}, y_{2,2}, \dots, y_{n,2}, y_{1,k}, \dots, y_{n,k})^T$, and model the data as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_d\mathbf{d} + \mathbf{e}. \tag{2.2}$$

Here $\boldsymbol{\beta}$ is a vector of fixed effects (which can include both continuous and class variables); \mathbf{a} represents the additive effects; \mathbf{d} the dominance effects; and \mathbf{e} the residual consisting of random environmental effects. \mathbf{X} , \mathbf{Z}_a and \mathbf{Z}_d are (known) incidence matrices. The residual, additive and dominance effects are assumed to follow the multivariate normal distributions

$$\begin{aligned} \mathbf{e} &\sim N(\mathbf{0}, \mathbf{V}_E \otimes \mathbf{I}_n), \quad \mathbf{a} \sim N(\mathbf{0}, \mathbf{G} \otimes \mathbf{A}) \quad \text{and} \\ \mathbf{d} &\sim N(\mathbf{0}, \mathbf{V}_D \otimes \mathbf{A}), \end{aligned} \tag{2.3}$$

where \otimes denotes outer product (see appendix B in the electronic supplementary material); \mathbf{I}_n is the $n \times n$ identity matrix; \mathbf{A} is the genetic additive relationship matrix; \mathbf{A} is the matrix of fraternity coefficients; and \mathbf{V}_E , \mathbf{G} and \mathbf{V}_D are the

$k \times k$ variance–covariance matrices of the environmental, additive and dominance effects to be estimated.

In appendix B in the electronic supplementary material, we describe how the posterior distribution of the model can be computed using either Gibbs or Metropolis–Hastings sampling. We implemented these procedures with MATHEMATICA v. 5.2 (Wolfram Research, Inc. 2005), the codes being available from the first author. In the Bayesian framework, priors need to be specified for $\boldsymbol{\beta}$, \mathbf{V}_E , \mathbf{G} and \mathbf{V}_D . In what follows, we assume a flat prior for $\boldsymbol{\beta}$, and either a flat or an inverse Wishart prior for the covariance matrices. For a discussion on the choice of the prior, see Gelman (2006).

(c) Comparison of variance–covariance matrices

A number of methods have been proposed for comparing matrices of genetic (co)variance among populations (Steppan et al. 2002; Blows et al. 2004). Among the most popular methods is common principal component analysis (CPC, Phillips & Arnold 1999), which examines the matrix structure in terms of eigenvalues and eigenvectors. To illustrate, let the 2×2 matrices \mathbf{G}_1 and \mathbf{G}_2 be estimated for populations 1 and 2, and assume, e.g. that the two matrices have identical first eigenvalues, and that the associated eigenvectors are perpendicular. In this case, one would conclude that the maximal amount of additive variation is the same in the two populations, but the direction of the variation in the phenotypic space is as different as it can be. However, the first eigenvalues do not tell the full story, and the support for this conclusion is actually dependent on the second

eigenvalues. If the second eigenvalues are much smaller than the first ones, the matrices are indeed very differently oriented. But if the second eigenvalues are only a little smaller than the first eigenvalues, the two matrices may actually be very similar. Hence, while the full set of eigenvalues and eigenvectors contains all the necessary information, its biological interpretation can be difficult, especially in case of high-dimensional matrices (Houle *et al.* 2002).

In the context of quantitative genetics, a variance–covariance matrix represents a multinormal distribution (or some other probability distribution), and the underlying probability distribution can be considered more fundamental than its representation through the matrix. We suggest that matrices of additive variance can be compared by asking how similar the underlying probability distributions are. To do so, a measure of distance $d(f, g)$ between two probability distributions is needed. We next derive one such metric by examining to what extent two probability distributions are distinguishable. Let $f(\mathbf{x})$ and $g(\mathbf{x})$ denote the densities of two probability distributions on \mathbf{R}^d , and let \mathbf{x} be a random sample drawn from either of the distributions. As the probability that \mathbf{x} would originate from g is $g(\mathbf{x})/(f(\mathbf{x}) + g(\mathbf{x}))$, the probability of assigning a random draw from f incorrectly to g is

$$q(f, g) = \int_{\mathbf{R}^d} \frac{g(\mathbf{x})}{f(\mathbf{x}) + g(\mathbf{x})} f(\mathbf{x}) d\mathbf{x}. \quad (2.4)$$

As $q(f, g)$ is symmetric, it describes also the probability of assigning a random draw from g incorrectly to f . We note that while performing the integration of equation (2.4) can be difficult in high-dimensional spaces, it is easy to approximate the integral through a Monte Carlo approach. Letting $\{\mathbf{x}_i\}_{i=1}^N$ be a random sample from f , $q(f, g)$ is approximated by the mean of $g(\mathbf{x}_i)/(f(\mathbf{x}_i) + g(\mathbf{x}_i))$ as $N \rightarrow \infty$.

If the probability distributions f and g are indistinguishable, it holds that $q(f, g) = 1/2$, whereas for completely distinguishable distributions $q(f, g) = 0$. Hence, we may define the distance between two probability distributions as

$$d(f, g) = \sqrt{1 - 2q(f, g)} = \left\{ \frac{1}{2} \int_{\mathbf{R}^d} \frac{[f(\mathbf{x}) - g(\mathbf{x})]^2}{f(\mathbf{x}) + g(\mathbf{x})} d\mathbf{x} \right\}^{1/2}. \quad (2.5)$$

We have taken the square root to ensure that d is a proper distance measure, i.e. that it satisfies the triangle inequality $d(f, g) \leq d(f, w) + d(w, g)$ for any probability distributions f , g and w (appendix C in the electronic supplementary material). We note that our measure d is actually a special case of distances known as *f-divergences* (see Gibbs & Su (2002) for a review of probability metrics). While the metric d has an intuitive probabilistic interpretation, other probability metrics could be used as well. For the sake of comparison, we consider also the standard L_2 metric

$$d_{L_2}(f, g) = \sqrt{\int_{\mathbf{R}^d} [f(\mathbf{x}) - g(\mathbf{x})]^2 d\mathbf{x}}. \quad (2.6)$$

As we estimate the \mathbf{G} matrix using a Bayesian approach, the uncertainty in the estimate is described through a posterior distribution. To examine the support of two populations having different \mathbf{G} matrices, we first draw two random samples from both populations. Let us denote the two samples from population A by \mathbf{G}_1^A and \mathbf{G}_2^A , and the samples from population B by \mathbf{G}_1^B and \mathbf{G}_2^B . We next calculate the difference

$$\begin{aligned} \Psi_d(A, B) = & [d(\mathbf{G}_1^A, \mathbf{G}_2^A) + d(\mathbf{G}_1^B, \mathbf{G}_2^B)] \\ & - [d(\mathbf{G}_1^A, \mathbf{G}_2^B) + d(\mathbf{G}_1^B, \mathbf{G}_2^A)], \end{aligned} \quad (2.7)$$

where the first term represents the distances of the matrices within populations, and the second term the same distance between populations. Hence, the statistical support for the two matrices being different (in the sense of distance d) can be evaluated by sampling repeatedly from the two posterior distributions to obtain $P(\Psi_d(A, B) < 0)$. As variance–covariance matrices and the underlying multinormal distributions (with zero mean) are equivalent, we have simplified above the notation by using d as the distance between two matrices rather than the distance between two probability distributions.

(d) *Generation of simulated data*

To examine the performance of the estimation schemes, we constructed a number of simulated datasets. We restrict the analyses here to the univariate case and illustrate the multivariate case in the context of real data (§2e). We consider the two contrasting study designs of the Siberian jay pedigree and the NC II breeding design (figure 1). With respect to the genetic architecture, we consider a two-case scenario in the context of fitness (high dominance ($=V_D$) to additive genetic ($=G$) variance ratio) versus non-fitness traits (low dominance to additive genetic variance ratio, Merilä & Sheldon (1999)). The idea behind this division is that traits closely linked to fitness, such as life history and sexually selected traits, are assumed to be under strong directional selection, whereas traits more remotely linked to fitness, such as morphological traits, are thought to be under weak directional or stabilizing selection. In the case of the fitness trait, we assume the parameter values $G=0.1$, $V_D=0.4$ and $V_E=0.5$, whereas in the case of the non-fitness trait, we assume the values $G=0.4$, $V_D=0.1$ and $V_E=0.5$, where V_E refers to environmental component of variance.

We generated 100 simulated datasets for both parameter sets and for both pedigree types. We estimated the parameters back for each dataset using the Bayesian and REML methods as described in appendix B in the electronic supplementary material. As we assumed a (known) zero mean and did not include any fixed effects, REML is identical to ML and coincides with the maximum posterior likelihood as we assume here a flat prior for the variance components. In case of the Siberian jay pedigree, we assumed that the phenotypic traits of all individuals had been measured, whereas in case of NC II we assumed that phenotypic measurements of the parents were not available. We assumed the NC II design to consist of 16 blocks consisting of four sires, four dams, and four offspring for each pair, leading to 64 offspring in each such block, and a total number of 1024 offspring. We examined the effect of sample size by estimating the parameters for subsets of n individuals with measured traits, with $n=64, 128, 256, 512$ and 1024. To examine if and how ignoring dominance variance biases estimates of additive variance, we estimated the parameters using both the full model and a model with additive and residual effects only.

(e) *Rana temporaria data*

To compare Bayesian and non-Bayesian inferences in case of real data, we reanalysed data on the common frog *Rana temporaria*. The data (described in Laurila *et al.* (2002)) consist of a set of NC II breeding designs on two frog populations originating from Kiruna (Northern Sweden) and Lund (Southern Sweden). The two populations live in contrasting environments and are separated by a distance of 1400 km. To examine a possible interaction between

environmental and additive effects, three desiccation treatments were applied at the tadpole phase: control (water level kept constant at 750 ml during the 28-day developmental phase); slow (water level gradually reduced from 750 to 240 ml); and fast (water level gradually reduced from 750 to 62 ml). Cano *et al.* (2004) used REML to estimate the \mathbf{G} matrices (with four traits: development time, body mass, body length and tail length; all log transformed) separately for each of the six datasets (two populations and three treatments) and compared the matrices using Flury's hierarchical method (Phillips & Arnold 1999). Following Cano *et al.* (2004), we account for additive, maternal and residual effects but ignore here the dominance effects. The fixed effects include the grand mean, the matrix (the crosses were performed in two 5×5 matrices for both populations) and the block (the tadpoles were placed in six shelves in the laboratory).

Each case consists in principle of 300 individuals (two matrices of five dams \times five sires with six offspring for each pair) but some of the data were missing due to unfertilized eggs. Furthermore, some trait values are clear outliers under the assumption of normality. If the outliers are due to measurement error, or due to failure with the treatment of the individual, they should be removed from the data. Another possibility is that the exceptional trait values are not outliers, but they represent natural variation, in which case a fat-tailed model would be more appropriate. To allow for comparison with Cano *et al.* (2004), we assumed the standard multi-normal model and analysed the data with and without the outliers. We considered a trait value an outlier if it was outside the $1-10^{-5}$ central probability density of the normal distribution determined by the data (after controlling for fixed effects). This led to sample size of 245 (with two outliers removed), 250 (three), and 248 (two) individuals for the control, slow and fast treatments of the Kiruna population, and 252 (six), 264 (six), 265 (five) individuals for the same treatments of the Lund population.

We assumed an inverse Wishart prior with six degrees of freedom and with the scale matrix $0.01 \mathbf{I}$, where \mathbf{I} is the 4×4 identity matrix. The posterior distributions were sampled using the Metropolis–Hastings algorithm described in appendix B in the electronic supplementary material.

3. RESULTS

(a) Fraternity coefficients

To examine whether the approximations based on equation (2.1) are likely to deviate substantially from the true values in case of a complex pedigree, we calculated the fraternity coefficients for the Siberian jay pedigree using both a simulation approach (appendix A in the electronic supplementary material) and equation (2.1). As illustrated in figure 1c, equation (2.1) gives a good approximation for a subset of pairs of individuals. However, there is another subset for which equation (2.1) gives roughly a twofold overestimate, suggesting high covariation of the events \mathbf{X} and \mathbf{Y} (see §2a for notation). For some pairs of individuals, equation (2.1) gives an underestimate (figure 1c), hence in these cases we must have $\Pr(\mathbf{X}_2|\mathbf{X}_1) > \Pr(\mathbf{X}_2)$.

The resolution to disentangle dominance effects from additive effects is determined by the amount of variation in fraternity coefficients not explained by variation in coancestry coefficients. In the case of the NC II design, this is solely due to the presence of full-sibs (with

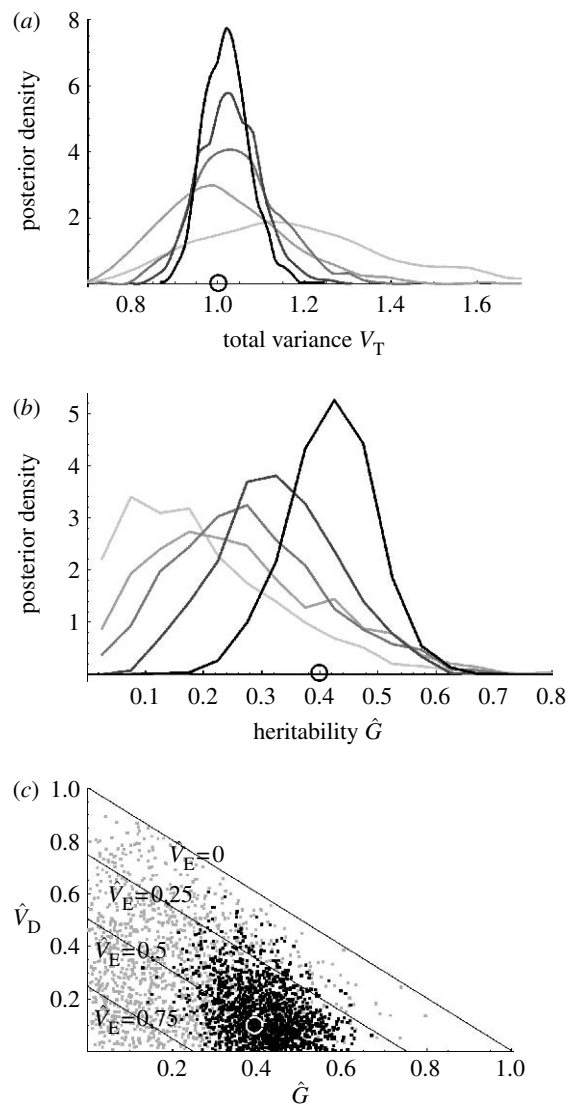


Figure 2. Examples of posterior distributions estimated back from computer-generated data. (a) The posterior distribution for the total variance $V_T = G + V_D + V_E$; the lines (from grey to black) corresponding to sample sizes $n = 64, 128, 256, 512$ and 1024 , and (b) the corresponding posteriors of heritability $\hat{G} = G/V_T$. (c) The joint posterior distribution of the relative variance components $\hat{G}_T, \hat{V}_D = V_D/V_T$ and $\hat{V}_E = V_E/V_T = 1 - \hat{G} - \hat{V}_D$. In (a–c), the circles show the true values used in data generation. The data was generated assuming the Siberian jay pedigree and parameter values corresponding to the non-fitness case.

$\Delta_{ij} = 0.25$), whereas in case of the Siberian jay pedigree the variation is of more or less continuous nature (figure 1d).

(b) Simulated data

As expected, the estimated posterior distributions concentrate around the correct values as the amount of data increases (figure 2). There is much less uncertainty in the total amount of variation (figure 2a) than in its division among the variance components (figure 2b,c). Hence, the joint posterior distribution over the variables \mathbf{G}, \mathbf{V}_D and \mathbf{V}_E is essentially concentrated on the plane $\mathbf{G} + \mathbf{V}_D + \mathbf{V}_E = c$, where c is a constant. We note that while the marginal posterior distribution of heritability is close to normally distributed with the larger datasets, this is not the case with

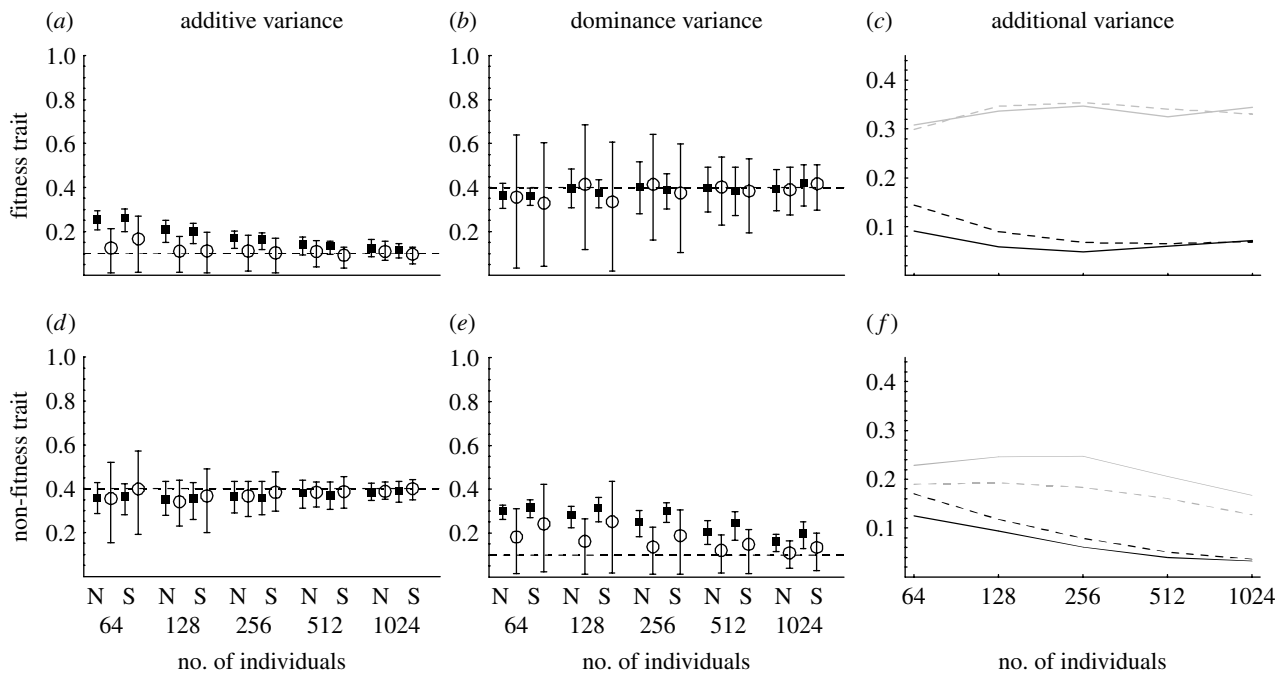


Figure 3. Performance of the Bayesian and REML estimation schemes. (a–d) The mean and 50% CI over the 100 replicated datasets of the posterior mean (box) and the REML estimate (circle) of the relative variance components $\hat{G} = G/V_T$ and $\hat{V}_D = V_D/V_T$. The letters N and S stand for the NC II and Siberian jay data structures, respectively. The dashed lines show the underlying true values. (e, f) The increase in the variance estimate (mean over the 100 replicated datasets of the posterior mean estimates) if the model used for parameter estimation does not include dominance variance. Black lines correspond to additive variance G , grey to residual variance V_E , continuous lines to Siberian Jay, and dashed lines to NC II data structures.

the small datasets (figure 2b). Hence, approximating uncertainty by assuming normality around the REML estimate can lead to a biased assessment in case of limited data.

Given the very contrasting natures of the Siberian jay pedigree and the NC II breeding design (figure 1), there is surprisingly little variation in the performance of the estimation schemes between these data structures (figure 3). Both data structures lead to almost identical estimates in case of the fitness trait. In case of the non-fitness trait, the estimates for both additive and dominance variances are somewhat larger for the Siberian jay pedigree than for the NC II breeding design, but even this difference is smaller than variation between replicated datasets. Consistent with the observation of Waldmann & Ericsson (2006), the REML estimates are less biased for small datasets than the posterior means. However, the REML estimates show much more variation among the replicated datasets than the posterior means. Another expected result is that the estimation of dominance effects is more difficult than that of the additive effects. Misztal (1997) argued that it requires of the order of 20 times more data to achieve the same accuracy for dominance variance as for additive variance, roughly consistently with the result of figure 3.

Ignoring dominance in the estimation process increased the estimates of both additive and residual variance (figure 3e, f). Interestingly, the increase in the estimated additive variance was almost identical between the fitness and non-fitness trait cases, although in the fitness trait case the ignored dominance component was four times greater than in the non-fitness trait case. There was no difference between the Siberian jay and NC II data structures in the overestimation of additive variance, but

in the non-fitness trait case there was a slight difference in the overestimation of residual variance.

(c) *Rana temporaria*

Variance–covariance matrices are positive definite and can be viewed as ellipsoids. Such a representation is adopted in figure 4, which illustrates the additive variance–covariance matrices as estimated for the six frog datasets (two populations and three treatments, with the outliers removed). As we considered four traits, the full matrices would correspond to four-dimensional ellipsoids (see appendix D in the electronic supplementary material), out of which figure 4 shows the marginal distribution restricted to two traits (body and tail lengths). More precisely, the ellipses depict for these two traits the 50% quantiles of the underlying multinormal distributions. Hence, the radii of the semi-axes are proportional to the square roots of the eigenvalues, and the directions of the semi-axes are determined by the eigenvectors. We note that in case of univariate analysis, the single eigenvalue would equal to the variance, and thus the radii of the semi-axes are proportional to standard deviations. Uncertainty in the \mathbf{G} matrix is not defined simply through confidence around individual matrix elements, but through the joint posterior distribution of the entire matrix (figure 4), including the dependence to other variance components (figure 2c). This allows one to examine matrix structure without the problem of estimation error being biased through possible nonlinear operations.

We will next examine whether the results from the Bayesian framework agree with those of Cano *et al.* (2004), who based matrix comparisons on REML estimates and the associated standard errors. As an example, Cano *et al.* (2004) found that in the control treatments, the genetic

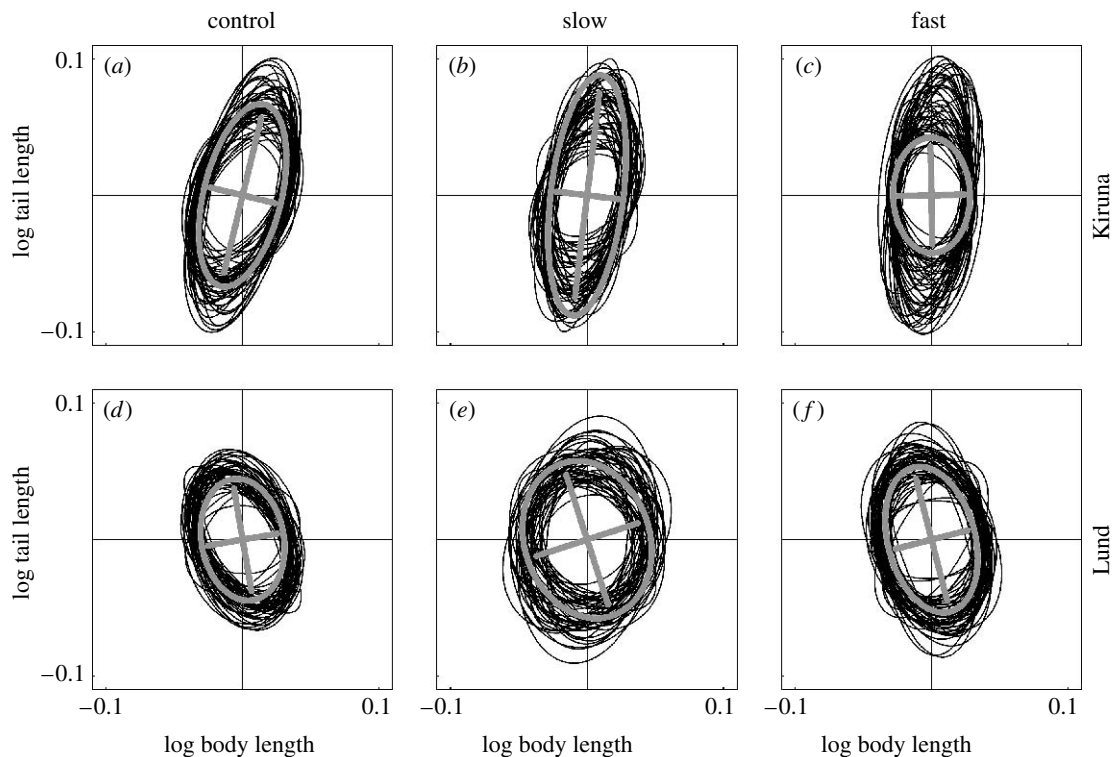


Figure 4. Posterior distributions of the \mathbf{G} matrices estimated for the three treatments of the two frog populations. (a–f) The figures show 2×2 marginal distributions of the full 4×4 \mathbf{G} matrices, obtained by restricting to the traits body length and tail length. (For all six combinations of trait-to-trait comparisons, see appendix D in the electronic supplementary material.) The ellipses represent the 50% quantiles of the underlying multinormal distribution. The grey ellipse with the semi-axes is the REML estimate and the black ellipses represent a random sample of 50 draws from the posterior distributions.

correlation between body and tail lengths was negative for the Lund population, but positive for the Kiruna population. The same result is apparent also in figure 4, in which the sign of the correlation is represented by the orientation of the ellipse. The evidence for a negative or positive correlation can be measured in terms of the posterior probability, computed from the posterior density for the genetic correlation by examining which fraction of the distribution is greater or smaller than zero. Denoting by ρ_P^T the correlation between body and tail lengths for population P ($P=L, K$) and treatment T ($T=c, s, f$), we obtain $P(\rho_K^c > 0) = 0.97$ and $P(\rho_L^c < 0) = 0.90$. Out of the 36 correlations (six datasets and six pairs of traits), Cano *et al.* (2004) reported 19 to be significant at the $p < 0.05$ level. Our Bayesian approach generally repeated the sign of the correlation, but we obtained high support (measured as $P(\rho > 0) > 0.975$ or $P(\rho > 0) < 0.025$) only for four correlations. The difference is not explained by the exclusion of the outliers, as including them did not change the results much, leading to this level of evidence for five correlations. The main explanation is that Bayesian posterior probabilities and classical p values relating to hypothesis testing are not equivalent, and hence cannot be directly compared. Cano *et al.* (2004) used the Hessian matrix to derive standard errors for the covariance components, which approach uses information on the shape of the likelihood surface locally around the REML estimate. In contrast, the posterior density is based on exploring the entire parameter space. As we can compute posterior distributions for derived quantities such as $\rho_L^c - \rho_K^c$, we can also assess directly the statistical support for statements such as $\rho_L^c < \rho_K^c$, giving $P(\rho_L^c < \rho_K^c) = 0.992$.

To examine differences among the six matrices in a more integrated manner, we summarize in table 1 various measures of matrix similarity. These include the two measures of distance between the underlying probability distributions (d and d_{L_2}), the leading eigenvalue of the matrix (λ , measuring the largest dimension of the ellipsoid), the ratio between the first two eigenvalues (ω , measuring the shape of the ellipse determined by the two largest eigenvalues), the angle between the leading eigenvectors (α , measuring difference in the orientation of the ellipsoid) and the volume of the ellipsoid (V , measuring the total amount of unconstrained variation). All of these measures, except the volume of the ellipsoid, indicated that the \mathbf{G} matrices of the two populations differ substantially if measured through the control treatment ($K-c$ versus $L-c$). The reason why the volume failed to indicate a difference is either because there is no such difference or maybe more likely because there is very limited information to estimate the smallest eigenvalues, and hence the prior may play a dominating role. The differences among the other pairs except $K-c$ versus $L-c$ are much less pronounced. While this does not rule out the possibility of identifying more subtle differences (e.g. in individual matrix elements), it illustrates that overall intrapopulation differences between treatments are so small that the data are not sufficient for their reliable quantification.

4. DISCUSSION

In this paper, we have presented a Bayesian framework for evolutionary quantitative genetics. We have examined how estimation techniques commonly used in animal

Table 1. Pair-wise comparisons between \mathbf{G} matrices estimated for the six datasets. (The datasets A and B are indexed as P - T , where $P=K$ (iruna), L (und) is the population and $T=c$ (ontrol), s (low), f (ast) is the treatment. The columns 3–6 represent the mean (upper row) and standard deviation (lower row) of the distance d (equation (2.5)), the associated statistical support ($P(\Psi_d(A, B) < 0)$) (equation (2.7)), and the same quantities based on distance d_{L_s} (equation (2.6)). V denotes the total volume of the ellipsoid representing the multinormal distribution (for a four-dimensional ellipsoid, $V=(1/2)\pi^2 r_1 r_2 r_3 r_4$, where r_i are the radii of the semi-axes). The leading eigenvalue is denoted by λ and the ratio between the second and first eigenvalues by ω . α denotes the mean angle $0 \leq \alpha \leq 90$ between the first and second eigenvectors, and Ψ_α the associated statistical support (calculated as in equation (2.7)).

A	B	d	Ψ_d	d_{L_s}	$\Psi_{d_{L_s}}$	$P(V_A < V_B)$	$P(\lambda_A < \lambda_B)$	$P(\omega_A < \omega_B)$	α	Ψ_α
$L-c$	$L-s$	0.60	0.18	261	0.17	0.84	0.25	0.48	59	0.23
		0.09		62						
$L-s$	$L-f$	0.58	0.32	227	0.37	0.32	0.22	0.66	53	0.33
		0.10		55						
$L-c$	$L-f$	0.53	0.38	237	0.33	0.71	0.12	0.67	58	0.41
		0.08		58						
$K-c$	$K-s$	0.55	0.30	234	0.28	0.26	0.70	0.55	24	0.34
		0.09		60						
$K-s$	$K-f$	0.54	0.34	250	0.36	0.34	0.28	0.65	30	0.38
		0.09		63						
$K-c$	$K-f$	0.60	0.13	271	0.11	0.15	0.45	0.70	36	0.14
		0.08		63						
$L-c$	$K-c$	0.66	0.02	297	0.02	0.80	0.99	0.08	65	0.04
		0.07		62						
$L-s$	$K-s$	0.62	0.14	263	0.11	0.23	0.74	0.17	41	0.29
		0.08		55						
$L-f$	$K-f$	0.61	0.11	278	0.09	0.24	0.73	0.17	59	0.23
		0.09		57						

breeding perform with evolutionary data, and illustrated how uncertainty in the \mathbf{G} matrix is quantified through the posterior distribution. We have shown that ignoring dominance effects in the animal model can lead to biased estimates of additive genetic variance, the effect being very similar in contrasting breeding designs. Based on the observation that (co)variance matrices represent a multinormal probability distribution, we have proposed that probability metrics provide a natural approach for comparing overall differences among populations. Finally, we found that approximations commonly used to compute fraternity coefficients can yield biased estimates in complex pedigrees, affecting the estimation of dominance variance.

To elaborate on the potential causes of biased fraternity coefficients, we first note that equation (2.1) mostly tended to give an overestimate (figure 1c), which is the expectation under inbreeding. As an extreme example of this type, consider the four parents of two individuals, and assume that all eight copies of a gene would be identical by descent. Then, the true fraternity coefficient for the two individuals would be one, but equation (2.1) would suggest a value of two, giving a twofold overestimate, and a number that is not a probability. In some cases, equation (2.1), however, predicted an underestimate. This may happen, e.g. if the pedigree has two lines, which are only little related to each other. As a hypothetical example, assume that all genes are identical by descent in all individuals of a line A (originating from a gene A in the ancestral population), and the same (with gene B) holds for a line B of individuals. Then, assume a single mating pair across the two lines, for which the offspring necessarily have one copy of gene A and one copy of gene B. Assume further that the lines A and B continue to be separated, except one mixed individual siring offspring

to line B. The gene A may or may not be lost in the subsequent generations of line B. Then consider, in the future generations, the fraternity coefficient between an individual in line A and an individual in line B. If the gene A was lost from line B, both events X_1 and X_2 have zero likelihood (for notation, see §2a). If gene A persisted (and invaded) in line B, both X_1 and X_2 can have a non-zero likelihood, leading to $\Pr(X_2|X_1) > \Pr(X_2)$. While the above examples are of hypothetical nature, the comparison based on the Siberian jay pedigree (with inbreeding coefficient up to 0.25) illustrates that the use of equation (2.1) can lead to substantial bias in case of long-term pedigrees.

In studies of natural populations, the estimation of the \mathbf{G} matrix has been difficult due to the complex hierarchical structure of quantitative genetic models. However, as illustrated here and earlier, especially in the context of animal breeding, Bayesian modelling provides a natural and flexible framework for the estimation of the \mathbf{G} matrix and other components of genetic architecture. The usage of the Bayesian approach provides a number of advantages over the traditional approaches, most importantly the ability to assess uncertainty in a robust way. In other words, we are also able to know what we do not know. This point becomes pronounced in the case of confounded variables. For example, it is well known that without the same parents breeding in different years, having multiple clutches per year, or cross-fostering, it is difficult to disentangle dominance effects from common environmental effects (Pante *et al.* 2002; Kruuk & Hadfield 2007). In our simulations, we included additive and dominance variance in the same model. As the fraternity coefficient correlates positively with the coancestry coefficient (figure 1d), the posterior distribution implied more certainty on the sum of the variances than on the individual variance components (figure 2). Whether the

larger uncertainty in the individual variance components matters or not depends on the further analyses the posterior is to be used for.

Fully accounting for uncertainty can make a major difference in applications, such as assessing the evolution of the \mathbf{G} matrix, or relative roles of selection and drift in phenotypic evolution (Merilä & Björklund 1999). If not assessed through the full joint probability distribution, error can propagate into quantities that depend on the estimate of \mathbf{G} , misleading the interpretations. This is likely to be the case especially in multidimensional problems, in which the accuracy of the evolutionary relevant characteristics of the \mathbf{G} matrix (such as orientation and dimensionality) depends nonlinearly on the individual matrix elements. Indeed, some approaches can lead to structural problems such as non-positive-definite (co)variance matrices, the treatment of which has called for additional ad hoc methods (e.g. bending; Hayes & Hill 1981).

We have developed here an integrated approach for matrix comparisons, which is based on examining how different the underlying multinormal distributions are with respect to a chosen probability metric. Imagine that genetic variability is represented by a hat, from which additive values for newborn individuals are drawn. Assume that there are two populations, both having their own hats. Then the distance d (equation (2.5)) between the two \mathbf{G} matrices can be viewed to measure if it is possible to tell from which hat the additive values of a given individual originate. Distance measures can be applied not only to \mathbf{G} matrices, but as well to other causal components, or, e.g. to the multivariate analogue of heritability, defined as $\mathbf{G}\mathbf{P}^{-1}$, where \mathbf{P} is the phenotypic covariance matrix. The strength and the weakness of the distance d is that it summarizes the difference between two \mathbf{G} matrices into a single number. To detail exactly in which way two matrices differ, this measure of matrix similarity can be supplemented with more detailed analyses of eigenstructure.

Studying the dimensionality of the \mathbf{G} matrix is important as it determines if and how \mathbf{G} constraints the pace and direction of evolution (Steppan *et al.* 2002). However, measuring dimensionality has proved to be difficult, different methods providing different answers, or the results depending on the degree of heritability (Hine & Blows 2006). While it is often possible to obtain good estimates of the leading eigenstructure (and there are direct methods to do so, see Kirkpatrick & Meyer (2004) and Meyer & Kirkpatrick (2005)), there is little statistical power to estimate the smallest eigenvalues, hence assessing matrix dimensionality calls for extensive datasets. This was illustrated by our results on matrix comparisons, in which the total volume of the ellipsoid failed to find any differences among the populations and treatments. However, consistent with the results of Cano *et al.* (2004), the Bayesian analysis found good evidence for differences in the leading aspects of shape (ω) and orientation (α) of \mathbf{G} between the two populations in the control treatment (table 1). The fact that the matrix comparisons based on the REML estimates found non-proportional differences in other population \times treatment combinations may be due to an inadequate assessment of the uncertainty (see appendix D in the electronic supplementary material) rather than real biological differences. This is not saying that the structures of the

matrices are necessarily the same, but that there is no conclusive information in the data.

We examined the statistical support for differences among \mathbf{G} matrices by comparing the posterior distributions within and between the populations. An alternative for comparing two \mathbf{G} matrices would be to parametrize two models, out of which the first one would contain a shared matrix for the two populations, and the second one a unique matrix for each population. In the context of matrix dimensionality, the competing models would be restricted to \mathbf{G} matrices of varying ranks. In the REML context, such competing models can be compared using likelihood-ratio tests (Shaw 1987). In the Bayesian framework, an analogous method would be to use Bayes factors or other means of measuring the distance of the data to a competing set of models (Gelman *et al.* 2004).

The reliability of the estimates of quantitative genetic parameters will be affected not only by the amount of data, but also by the underlying genetic architecture and the structure of the model used for its estimation. Our results confirm the fact that dominance variance is more difficult to estimate than additive variance (Miształ 1997), even without confounding effects common in evolutionary studies (Kruuk & Hadfield 2007). This makes it problematic to assess traits tightly linked to fitness, which are expected to have a higher dominance to variance ratio than non-fitness traits (Merilä & Sheldon 1999). Our simulations illustrated that ignoring dominance resulted in substantially biased estimates of additive genetic variance, irrespective of the underlying genetic architecture of the phenotypic trait. This is in contrast to Miształ *et al.* (1997), who found in the context of animal breeding (with data from 585 000 individuals) that almost all dominance variance became included in the residual variance. An additional problem with inbred populations is that the fraternity coefficients given by equation (2.1) can be biased, which can, however, be solved by computing them by the simulation approach proposed here.

We thank Elja Arjas, Bob O'Hara and Chaozhi Zheng for discussions on Bayesian estimation schemes and probability metrics, Jussi Alho and Sami Ojanen for help with preparing figure 1a and two anonymous referees for valuable comments. Thanks are also due to Bo-Göran Lilland for access to the Siberian jay data and Academy of Finland for financial support (grants 213457 and 211173 to O.O. and grants 118673 and 213491 to J.M.).

REFERENCES

- Arnold, S. J., Pfrender, M. E. & Jones, A. G. 2001 The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* **112**, 9–32. (doi:10.1023/A:1013373907708)
- Beaumont, M. A. & Rannala, B. 2004 The Bayesian revolution in genetics. *Nat. Rev. Genet.* **5**, 251–261. (doi:10.1038/nrg1318)
- Blasco, A. 2001 The Bayesian controversy in animal breeding. *J. Anim. Sci.* **79**, 2023–2046.
- Blows, M. W., Chenoweth, S. F. & Hine, E. 2004 Orientation of the genetic variance-covariance matrix and the fitness surface for multiple male sexually selected traits. *Am. Nat.* **163**, E329–E340. (doi:10.1086/381941)
- Cano, J. M., Laurila, A., Palo, J. & Merilä, J. 2004 Population differentiation in \mathbf{G} matrix structure due to natural selection in *Rana temporaria*. *Evolution* **58**, 2013–2020.

- de Boer, I. J. M. & van Arendonk, J. A. M. 1992 Prediction of additive and dominance effects in selected or unselected populations with inbreeding. *Theor. Appl. Genet.* **84**, 451–459. (doi:10.1007/BF00229506)
- Falconer, D. S. & Mackay, T. F. C. 1996 *Introduction to quantitative genetics*. Essex, UK: Longman.
- Gelman, A. 2006 Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **1**, 515–534.
- Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. 2004 *Bayesian data analysis*. Boca Raton, FL: Chapman and Hall/CRC.
- Gianola, D. & Fernando, R. L. 1986 Bayesian methods in animal breeding theory. *J. Anim. Sci.* **63**, 217–244.
- Gibbs, A. L. & Su, F. E. 2002 On choosing and bounding probability metrics. *Int. Stat. Rev.* **70**, 419–435.
- Hayes, J. F. & Hill, W. G. 1981 Modification of estimates of parameters in the construction of genetic selection indexes ('bending'). *Biometrics* **37**, 483–493. (doi:10.2307/2530561)
- Hine, E. & Blows, M. W. 2006 Determining the effective dimensionality of the genetic variance–covariance matrix. *Genetics* **173**, 1135–1144. (doi:10.1534/genetics.105.054627)
- Hofer, A. 1998 Variance component estimation in animal breeding: a review. *J. Anim. Breed. Genet. Zeitschrift Fur Tierzucht und Zuchtungsbiologie* **115**, 247–265.
- Houle, D., Mezey, J. & Galpern, P. 2002 Interpretation of the results of common principal components analyses. *Evolution* **56**, 433–440.
- Kearsey, M. J. & Pooni, H. S. 1996 *The genetical analysis of quantitative traits*. London, UK: Chapman and Hall.
- Kirkpatrick, M. & Meyer, K. 2004 Direct estimation of genetic principal components: simplified analysis of complex phenotypes. *Genetics* **168**, 2295–2306. (doi:10.1534/genetics.104.029181)
- Korsgaard, I. R., Lund, M. S., Sorensen, D., Gianola, D., Madsen, P. & Jensen, J. 2003 Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical and binary traits using Gibbs sampling. *Genet. Select. Evol.* **35**, 159–183. (doi:10.1051/gse:2003002)
- Kruuk, L. E. B. & Hadfield, J. D. 2007 How to separate genetic and environmental causes of similarity between relatives. *J. Evol. Biol.* **20**, 1890–1903. (doi:10.1111/j.1420-9101.2007.01377.x)
- Lande, R. 1979 Quantitative genetic-analysis of multivariate evolution. Applied to brain–body size allometry. *Evolution* **33**, 402–416. (doi:10.2307/2407630)
- Lande, R. & Arnold, S. J. 1983 The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226. (doi:10.2307/2408842)
- Laurila, A., Karttunen, S. & Merilä, J. 2002 Adaptive phenotypic plasticity and genetics of larval life histories in two *Rana temporaria* populations. *Evolution* **56**, 617–627.
- Lillandt, B. G., Bensch, S. & von Schantz, T. 2003 Family structure in the Siberian Jay as revealed by microsatellite analyses. *Condor* **105**, 505–514. (doi:10.1650/7117)
- Lynch, M. & Walsh, B. 1998 *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer.
- McGuigan, K. 2006 Studying phenotypic evolution using multivariate quantitative genetics. *Mol. Ecol.* **15**, 883–896.
- Merilä, J. & Björklund, M. 1999 Population divergence and morphometric integration in the greenfinch (*Carduelis chloris*)—evolution against the trajectory of least resistance? *J. Evol. Biol.* **12**, 103–112. (doi:10.1046/j.1420-9101.1999.00013.x)
- Merilä, J. & Björklund, M. 2004 Phenotypic integration as a constraint and adaptation. In *The ecology and evolution of complex phenotypes* (eds M. Pigliucci & K. Preston), pp. 107–129. Oxford, UK: Oxford University Press.
- Merilä, J. & Sheldon, B. C. 1999 Genetic architecture of fitness and nonfitness traits: empirical patterns and development of ideas. *Heredity* **83**, 103–109. (doi:10.1038/sj.hdy.6885850)
- Meyer, K. & Kirkpatrick, M. 2005 Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices. *Genet. Select. Evol.* **37**, 1–30. (doi:10.1051/gse:2004030)
- Mezey, J. G. & Houle, D. 2003 Comparing *G* matrices: are common principal components informative? *Genetics* **165**, 411–425.
- Misztal, I. 1997 Estimation of variance components with large-scale dominance models. *J. Dairy Sci.* **80**, 965–974.
- Misztal, I., Lawlor, T. J. & Fernando, R. L. 1997 Dominance models with method R for stature of Holsteins. *J. Dairy Sci.* **80**, 975–978.
- O'Hara, R. B. & Merilä, J. 2005 Bias and precision in *Q*(ST) estimates: problems and some solutions. *Genetics* **171**, 1331–1339. (doi:10.1534/genetics.105.044545)
- Pakkasmaa, S., Merilä, J. & O'Hara, R. B. 2003 Genetic and maternal effect influences on viability of common frog tadpoles under different environmental conditions. *Heredity* **91**, 117–124. (doi:10.1038/sj.hdy.6800289)
- Palo, J. U., O'Hara, R. B., Laugen, A. T., Laurila, A., Primmer, C. R. & Merilä, J. 2003 Latitudinal divergence of common frog (*Rana temporaria*) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. *Mol. Ecol.* **12**, 1963–1978. (doi:10.1046/j.1365-294X.2003.01865.x)
- Pante, M. J. R., Gjerde, B., McMillan, I. & Misztal, I. 2002 Estimation of additive and dominance genetic variances for body weight at harvest in rainbow trout, *Oncorhynchus mykiss*. *Aquaculture* **204**, 383–392. (doi:10.1016/S0044-8486(01)00825-0)
- Phillips, P. C. & Arnold, S. J. 1999 Hierarchical comparison of genetic variance–covariance matrices. I. Using the Flury hierarchy. *Evolution* **53**, 1506–1515. (doi:10.2307/2640896)
- Roff, D. A. 1997 *Evolutionary quantitative genetics*. New York, NY: Chapman and Hall.
- Schluter, D. 1996 Adaptive radiation along genetic lines of least resistance. *Evolution* **50**, 1766–1774. (doi:10.2307/2410734)
- Shaw, R. G. 1987 Maximum-likelihood approaches applied to quantitative genetics of natural populations. *Evolution* **41**, 812–826. (doi:10.2307/2408890)
- Shoemaker, J. S., Painter, I. S. & Weir, B. S. 1999 Bayesian statistics in genetics—a guide for the uninitiated. *Trends Genet.* **15**, 354–358. (doi:10.1016/S0168-9525(99)01751-5)
- Sorensen, D. & Gianola, D. 2002 *Likelihood, Bayesian, and MCMC methods in quantitative genetics*. New York, NY: Springer.
- Steppan, S. J., Phillips, P. C. & Houle, D. 2002 Comparative quantitative genetics: evolution of the *G* matrix. *Trends Ecol. Evol.* **17**, 320–327. (doi:10.1016/S0169-5347(02)02505-3)
- Waldmann, P. 2004 A quantitative genetic method for estimating developmental instability. *Evolution* **58**, 238–244.
- Waldmann, P. & Ericsson, T. 2006 Comparison of REML and Gibbs sampling estimates of multi-trait genetic parameters in Scots pine. *Theor. Appl. Genet.* **112**, 1441–1451. (doi:10.1007/s00122-006-0246-x)