**Quantifying individual variation in behaviour: mixed-effect modelling approaches**

Niels J. Dingemanse & Ned A. Dochtermann

*Journal of Animal Ecology*

doi: 10.1111/1365-2656.12013

–**Supplementary Material**–

For updates see: https://sites.google.com/site/howtoquantifybehaviour/

**Page**

**Contents 1**

**Supplementary Text S1** ERROR TERM DISTRIBUTIONS **3**

**Supplementary Text S2** EXAMPLES OF QUESTIONS ABOUT INDIVIDUAL VARIATION **3**

**Supplementary Text S3** HOW TO INCLUDE FIXED EFFECTS **4**

**Supplementary Text S4** INCLUDING ADDITIONAL RANDOM TERMS **4**

**Supplementary Text S5** THE EFFECTS OF DIFFERENCES IN PLASTICITY ON UNDERSTANDING

REPEATABILITY **5**

**Supplementary Text S6** IMPORTANT ASSUMPTIONS OF RANDOM REGRESSION MODELS **5**

**Supplementary Text S7** COMPARING VARIANCE COMPONENTS ACROSS DATASETS **6**

**Supplementary Text S8** CONTROLLING FOR OTHER LABILE ATTRIBUTES **6**

**Supplementary Text S9** FIXED EFFECTS THAT VARY WITHIN AND BETWEEN INDIVIDUALS **8**

**Supplementary Text S10** ESTIMATING COVARIANCES BETWEEN REACTION NORMS **10**

**Supplementary Text S11** ACCURACY OF CORRELATION ESTIMATES **11**

**Supplementary Text S12** POWER TO DETECT BETWEEN-INDIVIDUAL CORRELATIONS **12**

**Supplementary Text S13** ACCURACY OF REPEATABILITY ESTIMATES **13**

**Supplementary Text S14** POWER TO DETECT REPEATABILITY **13**

**Supplementary Text S15** CAUSES AND CONSEQUENCES OF WITHIN- AND

BETWEEN-INDIVIDUAL CORRELATIONS **14**

**Supplementary Text S16** TESTING HYPOTHESIZED COVARIANCE STRUCTURES **16**

**Supplementary Text S17** DO IT YOURSELF **18**

A. UNIVARIATE MMs WITH RANDOM INTERCEPTS FOR “INDIVIDUAL” **21**

B. BIVARIATE MMs FOR A SCENARIO WHERE TWO PHENOTYPIC ATTRIBUTES WERE BOTH ASSAYED REPEATEDLY AT THE SAME TIME (SCENARIO 3; TABLE 2) **26**

C. BIVARIATE MMs FOR A SCENARIO WHERE TWO PHENOTYPIC ATTRIBUTES WERE BOTH ASSAYED REPEATEDLY BUT NEVER AT THE SAME TIME (SCENARIO 4; TABLE 2) **33**

D. BIVARIATE MMs TO ESTIMATE REPEATABILITY FOR TWO DATASETS SIMULTANEOUSLY **39**

E. LIKELIHOOD RATIO TESTS **44**

EA. LRT-BASED SIGNIFICANCE OF IN UNIVARIATE MMs **45**

EB. LRT-BASED SIGNIFICANCE OF and IN BIVARIATE MMs FOR SCENARIO 5, TABLE 2 **48**

EC. LRT-BASED SIGNIFICANCE OF IN BIVARIATE MMs FOR SCENARIO 4,

TABLE 2 **52**

ED. LRT-BASED SIGNIFICANCE OF DIFFERENCES IN REPEATABILITY

BETWEEN TWO DATASETS **54**

**Supplementary Table S1**  **59**

**Supplementary Table S2**  **62**

**Supplementary Figure S1**  **64**

**Supplementary Figure S2**  **65**

**Supplementary References**  **67**

**Supplementary Text S1**

ERROR TERM DISTRIBUTIONS

MMs have been developed for both standard “linear” models in which error is expected to be normally distributed and to situations where it is not (e.g. binary, count, or proportional data). For these latter distributions, the MM is fit to data transformed by a “link function” and the errors are distributed according to a hypothetical distribution that is appropriate for the data type (Zuur et al. 2009). Although there are profound differences both in how these models are mathematically and numerically treated and how their fit to data is interpreted (Bolker et al. 2009), the modelling approaches discussed in this paper apply to normal and non-normal error distributions and thus our discussion is in the general context of MMs. However, the devil is in the detail. For example, Eqn. 2 in the main text does not apply to MMs with non-normal error distributions.

Specifically Eqn. 2 does not apply to calculating repeatabilities because of how the error terms are distributed (Eqn. 1a). As a result the denominator of Eqn. 2 is incorrect. Fortunately appropriate estimators of repeatabilities for non-normal errors are detailed elsewhere (Nakagawa & Schielzeth 2010). This issue with the error terms is also important when calculating covariances and correlations from multivariate models. While the between-individual covariances/correlations should be robust to these concerns, the within-individual covariances/correlations will not be. We are not aware of appropriate estimates of within-individual correlations for non-normal errors at this time.

**Supplementary Text S2**

EXAMPLES OF QUESTIONS ABOUT INDIVIDUAL VARIATION

We give five key examples of questions about individual variation in the main text. Here we give three further examples of questions about patterns of individual variation that are not addressed in main text but might be of interest to many animal ecologists:

1. *Is the average response of an individual correlated with its responsiveness (plasticity) to environmental change?* For example, do individuals that on average lay early in the season also show greater adjustments in lay date to changes in spring temperature? [Variance component 4 () in Table 1];
2. *Is an individual’s average response for one phenotypic attribute correlated with how plastic it is in another?* For example, is an individual that is on average relatively shy also relatively more plastic in how it adjusts its foraging behaviour to changes in perceived predation risk? [Variance component 7 () in Table 1];
3. *Is an individual’s level of responsiveness (plasticity) correlated across contexts?* For example, do individuals that show relatively pronounced adjustments in fat reserves in response to changes in mean resource availability also show relatively pronounced adjustments in fat reserves in response to changes in predictability of resource availability? [Variance component 8 () in Table 1].

**Supplementary Text S3**

HOW TO INCLUDE FIXED EFFECTS

In various sections of this paper, example models include fixed effects (i.e. *β*’s). The inclusion of fixed effects changes the interpretation of the parameters discussed above (see Kreft, Deleeuw & Aiken 1995; Enders & Tofighi 2007 for excellent discussions of this issue). For example, represents the grand mean value of average individual responses in models where no further fixed effects were fitted (Eqn. 1a). If fixed effect covariates were included, and *centred at the grand mean*, or expressed as deviations from individual mean values (*centred within an individual*), prior to inclusion in the model, would then represent the grand mean value of average individual responses when the fixed effect covariates are equal to zero. The choice to centre matters and whether and how centring should be applied depends on the question of interest (Kreft *et al.* 1995). For example, our description of as the expected average response of an individual is valid for Eqn. 1a and for models where all fixed effects were centred. Similarly, represents the between-individual variance at the position in phenotypic space where all fixed effects have the value zero (typically the reference category for categorical fixed effects); this is particularly relevant for random regression models discussed below. As a rule of thumb, clever centring can normally help provide meaningful zero points that raw covariates typically—though not always—lack (Enders & Tofighi 2007); we assume throughout our paper that fixed effect covariates were centred around their mean (of the population or individual for between- versus within-individual fixed effects, respectively), though we note that other types of centring may also be applied (Plewis 1989): left-centring may—for example—be applied to covariates with non-arbitrary zero values (e.g. time elapsed since the onset of an experiment).

**Supplementary Text S4**

INCLUDING ADDITIONAL RANDOM TERMS

When researchers are interested in whether individual repeatability has been inflated due to unmeasured habitat effects (which could occur when there is some level of repeatability in the location at which an individual is sampled—again, the norm in field studies), one could also include an additional random—instead of fixed—effect into the basic model given by Eqn. 1. For example, inclusion of random intercepts for territory (van de Crommenacker *et al.* 2011), or nest box identity (Browne *et al.* 2007), would enable the partitioning of phenotypic variance into between-individual, within-individual, and ‘among-habitat’ variance; repeatability could then be re-calculated (Eqn. 2) using the updated estimates of and derived from the extended model, and conclusions re-drawn. However, the biological interpretation of such analyses is complex as non-random distributions of individuals over habitats may represent a feature of the individuals’ typical phenotype. Translocations, where individuals are forced to settle at random, or experimental manipulations of environmental conditions, would be necessary to infer whether habitat and between-individual variances are distinct components.

**Supplementary Text S5**

THE EFFECTS OF DIFFERENCES IN PLASTICITY ON UNDERSTANDING REPEATABILITY

Importantly, inclusion of random slopes also allows for the evaluation of whether repeatability is constant versus a function of some environmental condition (when ≠ 0). By estimating repeatability is not assumed to be fixed but is instead explicitly allowed to vary over a gradient (), because the estimation of the intercept-slope covariance allows the individual variance to be a function of —the repeatability at a specific point on such a gradient is called *conditional repeatability* (Nakagawa & Schielzeth 2010). Fig. 1b illustrates a situation where the intercept-slope covariance is negative, showing greater between-individual variance in aggression (*y*) for the lower values of conspecific density (). Because Eqn. 5b explicitly assumes that the within-individual variance does not change with , repeatability of aggression would consequently decrease with increasing values of conspecific density for our worked example. In other words, in Fig. 1b, repeatability would be estimated as higher if sampling was conducted solely at low conspecific densities versus solely at high conspecific densities.

**Supplementary Text S6**

IMPORTANT ASSUMPTIONS OF RANDOM REGRESSION MODELS

Random regression models are extremely complex tools that can easily be misinterpreted. For example, behavioural ecology studies using random regression analyses often (e.g. all studies reviewed by van de Pol 2012) assume that within-individual variances (error) are homogeneous with regards to the fitted covariate (as does Eqn 5b; see also Cleasby & Nakagawa 2011 for how ecologists typically ignore heterogeneous errrors). In contrast, quantitative geneticists applying random regression analysis (Schaeffer 2004) typically both test and reject this assumption, see Brommer *et al.* (2008) and Dingemanse *et al.* (2012a) for, respectively, life-history and behavioural trait examples. These quantitative genetic studies therefore demonstrate that both between- and within-individual variation may be a function of the environment, and that key parameters (e.g. , and repeatability) may be mis-estimated when heterogeneous errors are not considered. Similarly, in cases where the fitted gradient itself shows *temporal autocorrelation* (e.g. gradual changes in population density over subsequent years), the model needs to include additional terms to appropriately estimate the within-individual variance (Pinheiro & Bates 2000; see Westneat *et al.* 2011 for a worked example).

**Supplementary Text S7**

COMPARING VARIANCE COMPONENTS ACROSS DATASETS

A particularly useful application of multivariate MMs is the statistical comparison of variance components (, ) across datasets. For example, one might be interested in whether males show greater levels of between-individual variance () in the same behaviour compared to female conspecifics (e.g. locomotor performance; Gilchrist 1996). Whenever repeated measures of the same behaviour are available for both female and male individuals, such questions can be answered by fitting a bivariate MM where behaviour of females is fitted as response variable *y* and the same behaviour of males is fitted as response variable *z*. The same random and fixed effect structure as in Eqn. 7 is used but the between-individual () and within-individual covariances () are now non-estimable (because we assume here that individuals cannot change sex), and therefore must be constrained to zero (Eqn. S1):

: (Eqn. S1)

:

The fit of this model can be compared with the fit of an alternative one in which and are constrained to be the same value (i.e. ), for example using a classic likelihood ratio test (Pinheiro & Bates 2000), though other methods may be more appropriate (Visscher 2006; Scheipl, Greven & Kuchenhoff 2008; Nakagawa & Schielzeth 2010). For example, when Bayesian approaches (i.e. MCMC-models) are applied, one could estimate variance components or proportions (i.e. repeatability) using sex-specific univariate MMs, and then interpret whether those are sex-specific by comparing the posterior distributions of the estimated point estimates (Hadfield 2010).

**Supplementary Text S8**

CONTROLLING FOR OTHER LABILE ATTRIBUTES

For asking questions about how much variation between or within individuals is shared with another labile phenotypic attribute, one might construct a bivariate MM (see section “**Multivariate MMs**” in the main text) where both phenotypic traits are treated as response variables (*y* and *z*), and where the covariance between individuals () and within individuals () are directly estimated. The between-individual covariance represents the covariance between the individual-mean values across the two phenotypic attributes, as might be caused by a genetic correlation or other effects that permanently affect the expression of both attributes. The within-individual covariance is the covariance between within-individual changes in expression between the attributes, as might be caused if measurement errors were correlated across the attributes, or because changes in one (metabolic rate) cause change in the other (behaviour) attribute within the individual. Whilst the covariance between traits is often of considerable interest (see section “**Multivariate MMs**” in the main text), if researchers are interested specifically in the variation of a behavioural response (or other labile trait) present independent of field metabolic rate, this variation can be estimated directly from a bivariate approach (Hansen et al. 2003). Specifically, the between-individual variance in response variable *y* that cannot be accounted for by the covariance with *z*, i.e. the “conditional” between-individual variance (), can be calculated as follows (Eqn. S2a):

(Eqn. S2a)

where and represent the between-individual variances in response variables *y* and *z*, respectively and is the between-individual covariance for *y* and *z*. Likewise, the between individual variance in *z* independent of *y* can be calculated as (Eqn. S2b):

(Eqn. S2b)

Similarly, the within-individual variance in response variable *y* not accounted for by the covariance with *z* () is given by the following equation (Eqn. S2c):

(Eqn. S2c)

where and represent the within-individual variance in response variable *y* and *z*, respectively. The conditional residual variance for *z* independent of *y* can be obtained in the same manner. As was the case for a simple univariate MM (e.g. Eqn. 4 in the main text), the effects of various fixed effects on can be estimated by comparing this variance component (and its within-individual counterpart) for models where certain fixed effects were included versus excluded. Furthermore, and can be used to calculate appropriate repeatability for *y* independent of *z*; e.g. the conditional repeatability of a behaviour independent of field metabolic rate, the question of initial interest. (Eqn. S2d):

(Eqn. S2d)

**Supplementary Text S9**

FIXED EFFECTS THAT VARY BOTH WITHIN AND BETWEEN INDIVIDUALS

Our earlier discussion of fixed effects focused on predictor variables that varied either between (B, ) or within individuals (W, ), or where one level was considered and another explicitly ignored (Eqn. 3). For cases where fixed effects vary *both* between and within individuals, relationships between response (*y*) and predictor variables (*x*) may also vary between these two hierarchal levels (illustrated in Fig. S1where the average value of the covariate varies between individuals). This variation between the two levels can occur when the between- and within-individual associations between phenotype (*y*) and environment (*x*) do not result from the same proximate mechanism (van Noordwijk & de Jong 1986; Reznick, Nunney & Tessier 2000). For example, in wild passerine birds individuals increase the speed of their exploratory behaviour (‘boldness’) from winter to spring, causing a *positive* relationship between behaviour (*y*) and time of year (*x*) *within* individuals (Dingemanse *et al.* 2012b). At the same time, relatively shy birds are often harder to capture (Biro & Dingemanse 2009), and might therefore be captured, on average, later in the season. This would lead to a *negative* relationship between behaviour and time of year *between* individuals. In this example, there are two proximate mechanisms affecting the relationship between phenotype and environment: phenotypic plasticity (within individuals) and differences in capture rates (between individuals). These contradictory effects can be teased apart using *within-group centring* techniques (Davis, Spaeth & Huson 1961; Raudenbush 1989b; Kreft *et al.* 1995; Snijders & Bosker 1999). In the context of variation between and within individuals, this method has been advocated in ecology under the term *within-subject centring* (van de Pol & Verhulst 2006; Snijders & Bosker 1999; van de Pol & Wright 2009). Here we discussed the approach detailed fully by van de Pol & Wright (2009), focus on the types of patterns of individual variation the approach addresses, and introduce concerns about its use.

When a continuous fixed effect (*x*) varies both between and within individuals, one can fit the following model (van de Pol & Wright 2009) (Eqn. S3):

(Eqn. S3)

Eqn. S3 initially appears equivalent to Eqn. 3. However, here we are modelling the response variable (*y*) not as a function of the individual-mean value for the continuous fixed effect (i.e. in Eqn. 3) but rather as a function of its specific value at instance *i*. Such an approach might be taken if we record the mass of an individual every time that we record the phenotypic attribute of interest. Since the range of masses experienced by individuals will vary, the dependence of on () is a mix of both between- and within-individual dependences, hence the subscript “B&W”. This conflation of between- and within-individual effects can be separated by calculating the mean covariate value for the individual (, as in Eqn. 3), and the deviation of the covariate from the mean for an individual for each measurement (). Both of these predictor variables (instead of just ) are then included in modelling the phenotypic response (Eqn. S4):

(Eqn. S4)

where represents the within-individual (W) dependence, and represents the between-individual (B) dependence of *y* on *x*. Using this approach the effects of *x* on *y* can be separated into those that are actually a function of *x* () and those that are a function of individuals being measured over different values of *x* ().

In natural populations inclusion of the between-individual component, , would help to avoid pseudo-repeatability (see above), whilst inclusion of the within-individual component, , would enable quantification of population-average phenotypic plasticity. Often it is , the population-average level of phenotypic plasticity, that is of key interest (van Noordwijk & de Jong 1986; van de Pol & Verhulst 2006)—necessitating that researchers more broadly use Eqn. S4.

There are a number of key applications for studying individual variation where within-individual subject centring might be useful. We focus in particular on the observation that different types of individuals are often non-randomly distributed over environments (e.g. due to *genotype by environment correlations*; Conover & Schultz 1995). This may occur when individuals either create (Laland, Odling-Smee & Feldman 1999; Vandermeer *et al.* 2004) or preferentially settle in environments where they perform well, as commonly suggested in the human psychology literature (Scarr & Mccartney 1983). For example, the aggressive individuals we discussed earlier might preferentially settle in areas with high conspecific densities (i.e. ) but might not alter their phenotype in response to changes in density (i.e. ), producing a pattern of variation where (illustrated in Fig. S1). In contrast, cases where are consistent with a scenario where all individuals share the same reaction norm, and where repeatable variation solely reflects pseudo-repeatability (Dingemanse *et al.* 2010c). For such a scenario, animals living in high-density populations would, on average, be more aggressive compared to those living in low-density populations (i.e. ). In other words, the inclusion of different terms as modeled by Eqn. S4 may provide quantitative evidence for whether individuals are non-randomly distributed over environments with respect to the intercept of their reaction norm (i.e. their average phenotype).

Though within-subject centring can reveal complex patterns of variation that would otherwise remain hidden, the method has also attracted criticism (Longford 1989; Plewis 1989; Raudenbush 1989a; Paccagnella 2006). For example, estimates derived for within-subject centred covariates (i.e. in Eqns. 6 and S4) may suffer reduced precision when the sample size (in terms of sampled individuals) is small (Plewis 1989), or they could be estimated with bias (Phillimore *et al.* 2010).

Examples given in the main text (section “**Estimating individual variation in plasticity**”) demonstrate how the usage of random regression analysis can provide key information about the stability of repeatability through the modelling of environment-dependent expression of between-individual variation. The partitioning of variance given above (Eqn. S4) between and within individuals can be taken further by allowing individuals to differ in their ability to respond to *x* by including the random slopes introduced earlier (i.e. Eqn. 5). Generalizing Eqn. S4 to a situation in which the plasticity of individuals varies produces (Eqn. S5):

(Eqn. S5)

This formulation reveals considerable information regarding how variation is partitioned within a population and is particularly important when asking questions about the presence of individual variation based on observations in natural populations where subject animals have control over the conditions under which they are assayed (Dingemanse *et al.* 2010c).

**Supplementary Text S10**

ESTIMATING COVARIANCES BETWEEN REACTION NORMS

Our discussion of multivariate MMs in the main text focused on the estimation of covariances between average levels of different phenotypic attributes (). Multivariate MMs can also be used to estimate (*i*) the covariance between the average level of one phenotypic attribute and the level of plasticity of another (), and (*ii*) covariances in levels of plasticity across two phenotypic attributes (). The former type of analysis might be conducted, for example, when asking whether within-individual changes in stress physiology covary with behavioural type (Dingemanse, Edelaar & Kempenaers 2010b; Koolhaas *et al.* 2010). The latter type of analysis would be appropriate when asking whether levels of plasticity are correlated across different phenotypic attributes (Dingemanse *et al.* 2010c; Sih & Bell 2008). For example, Biro et al. (2010) showed that individuals of the coral reef fish *Pomacentrus bankanensis* differed consistently in behavioural plasticity over a temperature gradient for both boldness and aggressiveness. BLUP-based estimates for individual level of plasticity in boldness were not correlated with those for aggressiveness, suggesting that plasticity in those phenotypic attributes were not underpinned by the same proximate mechanism. One can address such questions about covariance between reaction norm components of different phenotypic attributes whenever the same set of individuals is repeatedly assayed for two (or more) phenotypic attributes under a range of environmental conditions (covariate , e.g. conspecific density). Statistically this is done by fitting a bivariate version of the random regression model detailed in Eqns. 5 (Eqn. S6):

(Eqn. S6)

 

:

where the between-individual variance-covariance matrix contains four variance components: the variance in average level (intercept) of phenotypic attribute *y* () and *z* (), and the variance in phenotypic plasticity of *y* () and *z* (). This matrix also contains six corresponding covariance components: the covariance between average level (intercept) across the attributes (), the covariance between reaction norm intercept and slope for attributes *y* () and z (), the covariance between intercept of one attribute and slope of the other (, ), and the covariance between the slopes of the two attributes ().

**Supplementary Text S11**

ACCURACY OF CORRELATION ESTIMATES

To determine the accuracy of correlations at both the between- and within-individual level as derived from either single samples per individual or from mixed models we generated simulated datasets from known population distributions. Simulated datasets were generated according to Eqn. 7a by drawing values for and for each trait (*y* and *z*) from multivariate normal distributions as modeled by Eqn. 7b.

and for both *y* and *z* were set to 1, leading to population level repeatabilities of 0.5 for both traits. and were set at the population level according to values shown to the right of panels in Fig. 4a. 100 simulated datasets were generated for each combination of and for sample sizes of 25 through 150 individuals (in increments of 25) with 2, 4, 6, or 10 samples per individual. This resulted in a total of 9600 simulated datasets.

For each of these datasets we calculated the actual values of and since they will differ to some extent from the population values. Next, for each dataset, we then used mixed-models formulated according to Eqns. 7a and 7b to estimate between- and within-individual correlations. Estimation was conducted using the MCMCglmm library in *R* (v2.13; see Supplementary Text S17B below). We used inverse gamma priors, an MCMC sampling scheme of 13000 total iterations with a 3000 iteration burn-in and sampling (thinning) interval of 10. In trial runs these conditions led to satisfactory mixing and low levels of autocorrelation. We then compared the mixed-model estimates of the between- and within-individual correlations to the known parameters of the sample and quantified the accuracy of estimates as the root mean square error (RMSE) (Eqn. S7):

(Eqn. S7)

where *n* is the number of simulated datasets, *r* is the estimated correlation, and is the actual correlation (either between- or within-) for the sample. RMSE was calculated separately for each set of 100 datasets at each of the 96 combinations of number of individuals, samples per individual and and conditions. For each of the 9600 datasets we also calculated the correlation between *y* and *z* using a single sample per individual. We then calculated the accuracy with which this correlation estimated either the known between- or within-individual correlation as the RMSE.

**Supplementary Text S12**

POWER TO DETECT BETWEEN-INDIVIDUAL CORRELATIONS

To determine the power available to statistically detect between-individual correlations we again generated simulated datasets as described in Supplementary Text S11 according to Eqns. 7a and 7b. In contrast to the methods of Supplementary Text S11, for the purposes of determining power requirements we fixed to 0 at the population level and varied from 0.1 to 0.9 in increments of 0.2. 500 simulated datasets were generated for each increment of at sample sizes of 25 through 200 individuals (in increments of 25) with 2, 4, 6, or 10 samples per individual. This resulted in a total of 80,000 simulated datasets.

For each dataset, we then fit a mixed-model to estimate between-individual correlations according to Eqns. 7a and 7b. We then fit a second mixed model according to Eqn. 7a but with the following bivariate normally distributed variance-covariance structure (Eqn. S8):

: (Eqn. S8)

The two models were fit using the lme4 library in *R* (v2.13) using the syntax:

lmer(yz~-1+y\_rand+z\_rand+(y\_rand+z\_rand-1|ind))

for the first model in which *y* and *z* covarycorresponding to Eqns. 7a and 7b but with = 0; and the syntax:

lmer(yz~-1+y\_rand+z\_rand+(y\_rand-1|ind)+(z\_rand-1|ind))

for the second model, which fits the data to Eqns. 7a and S8. For both models “yz” corresponds to the observed phenotypic value, “y\_rand” and “z\_rand” is whether a particular value is for either trait *y* or *z*, and “ind” is a particular individual’s identifier.

Finally, for each dataset, we conducted a likelihood ratio test to determine whether the estimated correlation was statistically supported at the 0.05 level. The proportion of the 500 datasets at each parameter combination in which a significant correlation was obtained was then calculated, these proportions are plotted in Fig. 4b.

**Supplementary Text S13**

ACCURACY OF REPEATABILITY ESTIMATES

To determine the accuracy of repeatability estimates as derived from mixed models we generated simulated datasets from known population distributions. Simulated datasets were generated according to Eqn. 1a by drawing values for and from normal distributions as modelled by Eqns. 1a and 1b.

and were varied to produce repeatabilities ranging from 0.1 to 0.9 in increments of 0.2. 750 simulated datasets were generated for each level of repeatability for sample sizes of 25 through 150 individuals (in increments of 25) with 2, 4, 6, or 10 samples per individual. This resulted in a total of 120000 simulated datasets.

For each of these datasets we used mixed-effect models formulated according to Eqn. 1 to estimate between- and within-individual variances, from which repeatability was estimated using Eqn. 2. Estimation was conducted using the lme4 library in *R* (v2.13; see section A in Supplementary Text S14 below). We then compared the mixed-model estimates of repeatability to the generating value and quantified the accuracy of estimates as the root mean square error (RMSE) according to equation S6. RMSE was calculated separately for each set of 750 datasets at each of the 160 combinations of number of individuals, samples per individual and repeatability conditions.

**Supplementary Text S14**

POWER TO DETECT REPEATABILITY

To determine the power available to statistically detect repeatability we generated datasets as described in Supplementary Text S13 but generated 1100 datasets for each combination of sample sizes and repeatabilities for a total of 176000 datasets. For each dataset, as above, we fit a mixed-model to estimate between- and within-individual variances according to Eqns. 1a and 1b. We then fit a second model omitting a random factor. Following Martin *et al*. (2011) we fit the first model using the lme4 library in *R* (v2.13) and the syntax:

lmer(y~1+(1|ind))

As also done by Martin *et al*. (2011), we fit the second model using a simple linear model from the stats library in *R* with the syntax:

lm(y~1)

While the lme4 and stats libraries do not necessarily produce commensurate likelihoods we verified likelihoods for the second model versus a mixed-model including a dummy variable and the two had a correlation of 1. Details of *R* syntax are as above.

Finally, for each dataset, we conducted a likelihood ratio test to determine whether the estimated correlation was statistically supported at the 0.05 level. The proportion of the datasets at each parameter combination in which a significant correlation was obtained was then calculated, these proportions are plotted in Fig. 5b.

**Supplementary Text S15**

CAUSES AND CONSEQUENCES OF WITHIN- AND BETWEEN-INDIVIDUAL CORRELATIONS

What is the biological underpinning of between-individual and within-individual correlations?

Proximately, we can distinguish three main contributors to phenotypic correlations: genetic mechanisms, environmental mechanisms, and methodological artefacts (Fig. 3). Here, we discuss in detail how these mechanisms can shape correlations at different levels.

Genetic variation underpins phenotypic correlations whenever phenotypic attributes are linked through *genetic correlations*. This would be the case either if the same genes affect the expression of multiple phenotypic attributes (*pleiotropy*) or when genes affecting the expression of one phenotypic attribute are correlated with the genes affecting the expression of another (*linkage disequilibrium*) (Lynch & Walsh 1998). The presence of a genetic correlation implies that the evolutionary responses of phenotypic attributes are linked (Lande 1979), potentially constraining how populations respond to selection. In the context of behaviours, this potential for constraint generates the primary evolutionary implications of “behavioural syndromes” or “animal personalities” (Dochtermann & Roff 2010; Dochtermann 2011). Genes are attributes of individuals, and genetic variation thus contributes to variation at the between-individual level. Hence, genes that affect *multiple* aspects of the phenotype cause both between-individual (i.e. repeatable) variation in *y* and *z* () and between-individual correlations (Fig. 3).

Environmental variation can also underpin correlations through a variety of mechanisms, both between and within individuals (Fig. 3). At the between-individual level environmentally-induced correlations are called *permanent environment correlations*, at the within-individual level simply *environmental correlations* (Fig. 3). We discuss three key environmental mechanisms—both permanent and otherwise—of interest to ecologists. While these mechanisms are neither exhaustive nor mutually exclusive they illustrate why discriminating between between- and within-individual correlations is biologically relevant.

First, environmental effects can result in between-individual correlations through permanent environment correlations when the same environmental factor has relatively long-term effects on the expression of multiple phenotypic attributes. Examples of such environmental effects are the amount of parental investment or level of sibling competition that an individual experiences during ontogeny, or the amount of experience with a stimulus while learning (Stamps & Groothuis 2010). Environmental maternal and paternal effects (Lynch & Walsh 1998) can also be considered in this category (Fig. 3).

Second, there are environmental effects that can influence both between- *and* within-individual correlations. This depends on whether the environment varies at the between- or within-individual level. For example, in humans, individuals might wear shorts and go bare-footed on warms days but wear long trousers and shoes on cold days, resulting in a within-individual correlation due to variation in temperature experienced by an individual. At the same time, people living in the Bahamas typically experience higher temperatures, and therefore more often wear shorts and go bare-footed, compared to people living in Alaska; a between-individual correlation then emerges due to individual differences in average temperature. In other words, the same proximate mechanism (temperature) can cause correlations that are quantitatively similar at more than one level. This example illustrates that correlations between labile attributes observed through direct observations in natural populations confound individual and habitat variation, and therefore may lead to inappropriate inferences regarding ecological effects on individual variation (Dingemanse *et al.* 2010c).

A single environmental factor can also result in correlations with conflicting directions across different levels, cases where raw correlations provide limited information (as discussed in the section “**Utility of MMs versus alternatives**”). This occurs in situations where individuals have to trade-off either time or energy investment in multiple costly actions, a common concept in life-history research (Stearns 1992) and behavioural ecology (Westneat & Fox 2010), but individuals also differ in their access to or acquisition of energy. Using humans as an example again, money spent on a house cannot be spent on a car but wealthier individuals may be able to buy both big houses *and* big cars compared to poor people who can afford neither or only a small house and small car (van Noordwijk & de Jong 1986; Reznick *et al.* 2000). Hence house and car size may thus correlate *positively* between individuals and *negatively* within individuals. Here, the between-individual variation is again driven by permanent-environment variation, either because certain individuals acquired more resources early in life or because some individuals settled where resources were consistently plentiful. Meanwhile the within-individual correlation is driven by the trade-off in allocating money toward either a house or car. For this reason, life-history researchers interested in correlations between- versus within-individuals should aim to include a proxy for local environment into their models (e.g. random intercepts for the identity of the territory, nest box, etc.), and thereby clarify how local environment variation, between-individual variation, and within-individual variation contribute to raw phenotypic correlations (for an excellent introduction and example see Browne *et al.* 2007).

Third, different environmental factors—each influencing the expression of a single phenotypic attribute—may themselves be correlated, thereby inducing correlations either within or between individuals. For example, we can imagine that predator populations can only persist in habitats where prey densities are sufficiently high; between-individual correlations between phenotypic attributes of their prey may thus occur because the expression of one attribute is affected by conspecific density (e.g. aggression), whereas another one is affected by experience with predators (predator-inspection behaviour).

Finally, within-individual correlations in particular can also be underpinned by *correlated measurement errors* (Fig. 3) or result from other methodological practices (e.g. due to effects of order in which phenotypic assays are conducted; Dochtermann 2010; Stamps & Groothuis 2010). Fortunately, such biases can be quantified and statistically controlled. Consider for example a situation where behavioural data are taken from videos. To control for measurement error, the same observer could screen all videos (‘observations’) twice in a randomized order, and incorporate random intercepts for observation identity into the MM to partition the within-individual (co)variances into measurement error versus residual sources of within-individual (co)variation.

**Supplementary Text S16**

TESTING HYPOTHESIZED COVARIANCE STRUCTURES

Multivariate MMs can also be used to test specific hypotheses about how multiple (>2) labile phenotypic attributes covary. The pattern of covariances among phenotypic attributes, which has been termed *syndrome structure* (Dochtermann & Jenkins 2007; Dingemanse, Dochtermann & Wright 2010a), may differ given competing explanations for how or which phenotypic attributes should covary. Elsewhere we have described the use of confirmatory factor analysis implemented as structural equation modelling to evaluate syndrome structures (Dingemanse *et al.* 2010a) but many of the same questions can be evaluated directly within an MM framework, which may more properly partition covariances to the between- and within-individual levels. In an MM framework, competing hypotheses about syndrome structure—i.e. the structure of the covariance matrix—can be evaluated by allowing some covariances to be estimated while constraining others to be zero. For example, Dochtermann & Jenkins (2007) used confirmatory factor analysis to evaluate eight competing hypotheses of how four behavioural attributes covaried in Merriam’s kangaroo rats (*Dipodomys merriami*). The confirmatory factor models they considered can be approximated by placing specific constraints on the between-individual covariance matrix (). For four behavioural attributes, the linear equations of the multivariate MM are (Eqn. S9a):

(Eqn. S9a)

where *v, w, y,* and *z* correspond to seed caching, individual flexibility, aggression, and boldness, respectively, in the Dochtermann & Jenkins (2007) example. As we have seen in Eqn. 7b, the random intercepts are modeled as having means of zero and distributed according to a multivariate normal distribution with a variance-covariance structure ():

 (Eqn. S9b)

The within-individual effects () would be modeled by a similar extension of Eqn. 8b. As an example, we highlight here three competing hypotheses for how these phenotypic attributes are expected to covary. First, previous literature might suggest that phenotypic attributes *v* and *w* (seed caching and individual variability) covary, and that *y* and *z* (aggression and boldness) covary due to genetic or permanent-environment effects (Fig. S2). This hypothesis would equate to a syndrome structure as presented in Fig. S2a which corresponds to a Ω*ind* of (Eqn. S9c):

 (Eqn. S9c)

By constraining specific elements of the covariance matrix to zero we have forced the pattern of relationships illustrated in Fig. S2a. Alternatively, three of the four phenotypic attributes (*v, w,* and *z*) might be hypothesised to covary. This pattern would be represented as a syndrome with a structure illustrated in Fig.S2b, which corresponds to a Ω*ind* of (Eqn. S9d):

 (Eqn. S9d)

In Fig. S2b (and S2c below) we represent the relationships amongst phenotypic attributes as being due to connections stemming from a shared underlying cause. The expected pattern of covariances for this relationship corresponds to Eqn. S9d but could also be modelled as being due to a latent variable (Dingemanse *et al.* 2010a; Dochtermann & Jenkins 2007). Finally, all four phenotypic attributes might be expected to covary due to genetic or permanent-environment effects (Fig. S2c), corresponding to a full syndrome structure estimated by the covariance matrix for Eqn. S9b. Assuming that these three hypothesized syndrome structures share the same hypothesized within-individual covariance structure, the fit of these models could then be statistically assessed using simple likelihood ratio tests for nested models (see above) or using information criteria (Garamszegi *et al.* 2009) to determine which of the hypothesized structures best correspond to the data. Interacting components of the phenotype may also be modelled, which would require more complex manipulations of Ω*ind* (e.g. Rosa *et al.* 2011).

**Supplementary Text S17**

DO IT YOURSELF

Here we provide specific programming code for fitting varies types of univariate and multivariate mixed-effect models (see Glossary). We further provide specific programming code for calculating likelihood ratio tests and differences in information criteria.

We detail how such analyses may be done for three commonly used software packages (*ASReml* v3.0, *R* v2.13, and SAS 9.2). All three packages are widely available, ASReml v3.0 is free only for university employees, *R* is freely available, and SAS must be purchased, but comes in student editions and is often available via institutional licenses. All three provide flexible programming options necessary for fitting and constraining complex MMs. We base our worked examples on simulated data (drawn from a bivariate normal distribution) available as Supplementary Information (Dataset S1-S3), which readers may use to learn how to address these specific questions before applying the advocated approaches to their own data.

DATASET DESCRIPTIONS

*DataS1.txt* represents an example where two phenotypic attributes, *y* and *z*, were both assayed repeatedly at the same time as in scenario 3/5 of Table 2 of the main text. Specifically, each of 50 simulated individuals (“*Indiv*”) was assayed five times for both attributes (i.e. 250 rows). A single predictor covariate, *“x”*, was assayed for each record. Following examples in the main text, *x* varied both within and between individuals, and we therefore also calculated both the mean covariate value for the individual (“*x\_avg*”, i.e. in Eqn. 6 of the main text), and the deviation of the covariate from the mean for an individual for each measurement (“*x\_dev*”; i.e. in Eqn. 6 of the main text). MMs often experience fewer problems with convergence when predictor variables are centred prior to analyses (e.g. Rasbash *et al.* 2005), whilst between-individual variances are normally measured for mean-centred predictor values (Wilson *et al.* 2010), and we therefore also extracted the population mean value of *x* from the raw values of *x* (“*x\_cen*”), and similarly extracted the population mean value of *x\_avg* from the raw values of *x\_avg* (“*x\_avg\_cen*”). In other words, the variables *x\_cen*, *x\_dev*, and *x\_avg\_cen* all have a mean value equal to zero. We also expressed the phenotypic attributes *y* and *z* in standard deviation units (*y\_std*, *z\_std*); these standardized variables thus both have a mean value equal to zero and a variance equal to one.

*DataS2.txt* represents an example where two phenotypic attributes, *y* and *z*, were both assayed repeatedly but never at the same time as in scenario 4 of Table 2 of the main text. Specifically, each of 50 simulated individuals was assayed five times for each attribute (e.g. *y* repeatedly in summer, and *z* repeatedly in winter). Because the attributes were never assayed at the same time, *DataS2.txt* holds twice as many rows compared to *DataS1.txt* (i.e. 500 instead of 250). “NA” is printed for missing data. We again expressed the phenotypic attributes *y* and *z* also in standard deviation units (*y\_std*, *z\_std*).

*DataS3.txt* represents an example where the same phenotypic attribute was measured repeatedly both for females (*y*) and males (*z*), five times for each individual, where individual numbers 1-50 are females, and individual numbers 51-100 are males. We again expressed the phenotypic attributes *y* and *z* also in standard deviation units (*y\_std*, *z\_std*).

INTRODUCTION TO THE DISCUSSED SOFTWARE PACKAGES

*ASReml* is a software package that provides a flexible means for a diverse array of univariate and multivariate MMs with normal errors, permitting (*i*) the estimation of variances, covariances, correlations, and ratios (e.g. repeatabilities) with their associated standard errors, and (*ii*) manipulation (e.g. constraining) of variance-covariance matrix elements (Gilmour *et al.* 2006). *ASReml* is commonly used in the field of quantitative genetics, as it allows for fitting a specific class of MMs –called “animal models”– that can partition variance components into additive genetic and residual components for pedigreed datasets (Wilson *et al.* 2010). Those interested in using *ASReml* are strongly advised to consult the WAMWiki ([http://www.wildanimalmodels.org](http://www.wildanimalmodels.org/)), which holds detailed background on how to get started with, and construct, both simple and complex MMs. Here we detail particularities of this software package that will help readers to get started with our worked examples. Specifically, there are four types of file that ASReml either uses or produces that feature in our worked examples. First, the **.as** file both calls and holds a description of the data file (**.dat**; tab-delimited) and contains the programming code for fitting a specific model. Second, the **.asr** file reports diagnostics of the analysis, including information on estimated variance components, significance of fixed effects, cautions regarding problems with convergence, and model fit (LogLikelihood). Third, the **.sln** file reports the fixed effect estimates and BLUPs for all levels of random effects, with their approximate standard errors. Fourth, the **.pin** file uses the estimated variance-covariance matrices to calculate user-specified estimates (with associated standard errors), which it reports in the **\*.sln** file. In our description of how to do MMs with ASreml, we solely detail programming code needed to construct specific models (i.e. **\*.as** files) and extract parameters of interest ( **\*.pin** files).

*R* is an open source programming language with a very active user base and it is these users that provide and develop the statistical libraries that are available. A variety of software packages are available for both linear and non-linear mixed effects models (reviewed by Bolker *et al.* 2009). While the majority of these packages were not developed specifically for the decomposition of phenotypic variance components by evolutionary ecologists, many are useful for such purposes. Here we provide programming code for univariate models for two libraries: lme4 and MCMCglmm**,** and multivariate code for the MCMCglmm library. The former library fits models using a restricted estimate maximum likelihood approach (other fitting options are available) while the latter uses a prior-informed Markov Chain Monte Carlo approach (i.e. a Bayesian approach). While both approaches require considerable care regarding model specification the latter also requires consideration of how the prior distribution of variance components is modelled. This is a difficult issue that is discussed by Hadfield (2010), the author of the MCMCglmm library and throughout the primary literature on Bayesian analyses. We thus recommend that users investigate that literature on their own.

Notably, programming specifications for both *ASReml* and the MCMCglmm *R* library were also provided by Wilson *et al.* (2010) in their discussion of models for the estimation of quantitative genetic parameters.

SAS, or “Statistical Analysis System”, was originally designed for agricultural statistics. The statistical package SAS 9.2 (the newest version, SAS 9.3, may not be available everywhere) contains a module called “PROC MIXED” which can flexibly be used to analyse a wide variety of mixed models. An excellent reference for many uses of PROC MIXED is Littell et al. (2006). An initial benefit of SAS 9.2 is that many elements of using it are user-friendly, including a point and click module called “Analyst” (no longer available in version 9.3), yet it retains flexibility through the use of a programming language. We provide some solutions to all the analyses described here, but SAS experts will surely know more elegant ones.

SAS uses its own data file, and SAS data files are housed in libraries, which can be located in any folder you choose and named to represent the project. Here we named the library “HowTo”, with the same file name as provided (e.g., HowTo.DataS1). Date files can be manipulated within the SAS environment, but there is a menu-driven procedure for uploading data files from MS Excel if the user prefers entering and modifying the data in that environment. What follows is the code necessary to convert a dataset in Excel to a SAS.dat file.

**PROC** **IMPORT** OUT= HOWTO.DATAS1

DATAFILE= “*YOUR LOCATION*\Datasets S1-S3.xls"

DBMS=EXCEL REPLACE;

RANGE="'Dataset S1$'";

GETNAMES=YES;

MIXED=NO;

SCANTEXT=YES;

USEDATE=YES;

SCANTIME=YES;

**RUN**;

Once this is completed successfully, all subsequent programs use the HOWTO library and the dataset name formatted as “HOWTO.*Dataset name*” to retrieve the data file.

A. UNIVARIATE MMs WITH RANDOM INTERCEPTS FOR “INDIVIDUAL”

*Application*: The example is for running a univariate MM with random intercepts for the factor “Individual”. The within- and between-individual effects of a single covariate (*x*) that varies both within and between individuals are estimated by fitting *x\_avg\_cen* (i.e. ) and *x\_dev* (i.e. ) as fixed effects into the model. , , , and repeatability, with their associated confidence (standard errors / credible intervals) are calculated. The example uses *DataS1.txt*; the phenotypic attribute *y* is used as the response variable. Note that inclusion of these fixed effects produces *conditional* values of , , , and repeatability. Furthermore, comparison of these values with values derived from an alternative model where these fixed effects were not included would allow for the (*i*) calculation of *unconditional* values, and (*ii*) the calculation of variance explained by the fixed effects between and within individuals.

*Specified model: Linear Eqn. S4 and random effect structure Eqn. 1b:*

:

:

*ASReml coding, output, and interpretation (5 parts)*

Part 1: coding of the .as file

**ASREML analysis**

**Indiv !A**

**x**

**x\_avg**

**x\_dev**

**y**

**z**

**x\_cen**

**x\_avg\_cen**

**y\_std**

**z\_std**

**C:\Howtodo\DataS1.txt !SKIP 1**

**!MVREMOVE**

**!MAXIT 5000**

**!SLOW 3**

**!EXTRA 5**

**!CONTINUE**

**y ~ mu x\_avg\_cen x\_dev !r Indiv**

**1 1 1 !STEP 0.001**

**0 0 ID 0.1 !S2==1**

**Indiv 1**

**Indiv 0 ID 0.1 !GP**

Part 2: coding of the .pin file

**# Pin file for (conditional) repeatability of a single phenotypic attribute**

**# (Co)variance components given in output (asr-file) are**

**# 1. trait A error variance**

**# 2. trait A Indiv variance**

**F PhenVarA 1+2 # 3. Phenotypic variance of A with SE**

**F ResVarA 1 # 1. Residual variance of A with SE**

**F IndVarA 2 # 2. Individual variance of A with SE**

**H RepeatA 2 3 # Repeatability of A with SE**

Part 3: Key output reported in the .asr file

**36 LogL=-229.472 S2= 1.0000 247 df 1.000 1.708 1.049**

**Final parameter values 1.0000 1.7079 1.0486**

**- - - Results from analysis of y - - -**

**Source Model terms Gamma Component Comp/SE % C**

**Residual identity 250 1.70789 1.70789 9.97 0 U**

**Indiv identity 50 1.04857 1.04857 3.67 0 P**

**Wald F statistics**

**Source of Variation NumDF DenDF F-inc P-inc**

**11 mu 1 48.0 12.54 <.001**

**8 x\_avg\_cen 1 48.0 0.43 0.518**

**4 x\_dev 1 199.0 0.60 0.439**

**Solution Standard Error T-value T-prev**

**4 x\_dev**

**1 0.451943 0.584447 0.77**

**8 x\_avg\_cen**

**1 0.125671 0.192746 0.65**

**11 mu**

**1 -0.590446 0.166743 -3.54**

**1 Indiv 50 effects fitted**

**Finished: 02 Mar 2011 08:51:48.784 LogL Converged**

Part 4: Variance components reported in the .pvc file

**3 PhenVarA 1 2.7565 0.31509**

**4 ResVarA 1 1.7079 0.17122**

**5 IndVarA 2 1.0486 0.28582**

**RepeatA = Indiv 2/PhenVarA 3= 0.3804 0.0711**

**Notice: The parameter estimates are followed by**

**their approximate standard errors.**

Part 5: Overview of parameter estimates (and where to find them)

**±SE = -0.590446±0.166743 (asr file)**

**±SE = 0.451943±0.584447, F1,199 = 0.60, P = 0.439 (asr file)**

**±SE = 0.125671±0.192746, F1,48 = 0.43, P = 0.518 (asr file)**

**±SE = 2.7565±0.31509 (pvc file)**

**±SE = 1.0486±0.28582 (pvc file)\***

**±SE = 1.7079±0.17122 (pvc file)**

**Conditional repeatability±SE = 0.3804±0.0711 (pvc file)\***

**Model Loglikelihood = -229.472 (asr file).**

**\*see section F for how to calculate significance using a likelihood ratio test**

*R coding, output, and interpretation (4 parts)*

Part 1: coding for MCMCglmm(Hadfield 2010) and lmer{lme4} R libraries

#MCMC fitted model using the MCMCglmm library

#uses the default priors and default iterations

#in most cases neither of these defaults are likely to be

#appropriate. Please see Hadfield (2010) for further

#discussion of both issues.

y.mcmc<-MCMCglmm(y~x\_avg\_cen+x\_dev,

random=~indiv, family="gaussian",

data=set1.data,verbose=FALSE)

#REML fitted model using the lme4 library

y.lme<-lmer(y~x\_avg\_cen+x\_dev+(1|indiv),data=set1.data)

Part 2: key output for both fitting approaches from summary(y.mcmc)

DIC: 885.4448

G-structure: ~indiv

post.mean l-95% CI u-95% CI eff.samp

indiv 1.081 0.4786 1.68 1000

R-structure: ~units

post.mean l-95% CI u-95% CI eff.samp

units 1.721 1.375 2.022 1000

Location effects: y ~ x\_avg\_cen + x\_dev

post.mean l-95% CI u-95% CI eff.samp pMCMC

(Intercept) -0.5841 -0.9453 -0.2616 1000 0.006\*\*

x\_avg\_cen 0.1273 -0.2229 0.5079 1000 0.490

x\_dev 0.4300 -0.6162 1.6845 1000 0.472

from summary(y.lme):

AIC BIC logLik deviance REMLdev

922.9 940.5 -456.5 910.4 912.9

Random effects:

Groups Name Variance Std.Dev.

indiv (Intercept) 1.0486 1.0240

Residual 1.7079 1.3069

Number of obs: 250, groups: indiv, 50

Fixed effects:

Estimate Std. Error t value

(Intercept) -0.5905 0.1667 -3.541

x\_avg\_cen 0.1256 0.1927 0.652

x\_dev 0.4518 0.5844 0.773

Part 3: calculating repeatability from lme and MCMC output

#repeatability can be calculated for each MCMC sample

#and the posterior mode and its

#95% credibility interval then extracted:

rep.y<-(y.mcmc$VCV[,"indiv"]/

(y.mcmc$VCV[,"indiv"]+y.mcmc$VCV[,"units"]))

posterior.mode(rep.y) #repeatability estimate

HPDinterval(rep.y) #95% credibility interval

Part 4: Overview of parameter estimates (and how to retrieve them)

from MCMCglmm (posterior modes between credibility intervals):

: -0.945 < -0.583 < -0.262 (posterior.mode(y.mcmc$Sol); HPDinterval(y.mcmc$Sol))

: -0.616 < 0.389 < 1.685 (posterior.mode(y.mcmc$Sol); HPDinterval(y.mcmc$Sol))

: -0.223 < 0.131 < 0.508 (posterior.mode(y.mcmc$Sol); HPDinterval(y.mcmc$Sol))

: 0.479 < 0.939 < 1.680 (posterior.mode(y.mcmc$VCV); HPDinterval(y.mcmc$VCV))

: 1.375 < 1.634 < 2.022 (posterior.mode(y.mcmc$VCV); HPDinterval(y.mcmc$VCV))

Conditional repeatability: 0.244 < 0.391 < 0.521 (described in Part 3)

from lme:

±SE = -0.5905±0.1667 (summary(y.lme))

±SE = 0.4518±0.5844 (summary(y.lme))

±SE = 0.1256±0.1927 (summary(y.lme))

= 1.0486 (summary(y.lme))

= 1.7079 (summary(y.lme))

Conditional repeatability = 0.3804 (eqn 2 and the above two values)

*SAS 9.2 coding, output, and interpretation (2 parts)*

Part 1: Model Statement

**proc** **mixed** data=HowTO.DataS1 method=reml cl alpha=**.05** covtest;

class INDIV;

model Y = X\_AVG\_CEN X\_DEV / htype=**3** ddfm=kr solution cl alpha=**.05**;

random int/subject=INDIV;

**run**;

Part 2: Output

SAS outputs data in several ways, with the default being text placed in an output window. The user can copy and paste this into other programs, but since it is text it is difficult to extract particular results except by selecting and copying. One can use the ODS system in SAS to output results directly into a spreadsheet program, but here we provide the text version. We have indicated the relevant parameters with inserted comments below:

Covariance Parameter Estimates

Standard Z

Cov Parm Subject Estimate Error Value Pr > Z Alpha Lower Upper

Intercept Indiv 1.0486 0.2858 3.67 0.0001 0.05 0.6550 1.9448

Residual 1.7079 0.1712 9.97 <.0001 0.05 1.4164 2.1002

[The “estimate” for intercept is , with “error” its standard error, and “lower” and “upper” indicating confidence limits.The “estimate” for residual is]

Fit Statistics

-2 Res Log Likelihood 912.9

AIC (smaller is better) 916.9

AICC (smaller is better) 916.9

BIC (smaller is better) 920.7

Solution for Fixed Effects

Standard

Effect Estimate Error DF t Value Pr > |t| Alpha Lower Upper

Intercept -0.5904 0.1667 48 -3.54 0.0009 0.05 -0.9257 -0.2552

x\_avg\_cen 0.1257 0.1927 48 0.65 0.5175 0.05 -0.2619 0.5132

x\_dev 0.4519 0.5844 199 0.77 0.4403 0.05 -0.7006 1.6044

[The “estimates” above are in order from top to bottom, , , and .]

Type 3 Tests of Fixed Effects

Num Den

Effect DF DF F Value Pr > F

x\_avg\_cen 1 48 0.43 0.5175

x\_dev 1 199 0.60 0.4403

Conditional repeatability may be calculated by hand by filling in the estimated variance components in Eqn. 2; we hope to provide information on the calculation of repeatability and its uncertainty (standard errors) in updated versions of this text on our website.

B. BIVARIATE MMs FOR A SCENARIO WHERE TWO PHENOTYPIC ATTRIBUTES WERE BOTH ASSAYED REPEATEDLY AT THE SAME TIME (SCENARIO 3; TABLE 2)

*Application*: The example is for running a bivariate MM with random intercepts for the factor “Individual”. , , , and repeatability are calculated for *y* and *z*, as well as their covariances (, , ) and correlations (, , ), with their associated confidence (standard errors / credible intervals). The example uses *DataS1.txt*; the phenotypic attributes *y* and *z* are used as the response variables.

*Specified model: Linear Eqn. 7a and random effect structure Eqn. 7b:*

:

:

*ASReml coding, output, and interpretation (5 parts)*

Part 1: coding of the .as file

**ASREML analysis**

**Indiv !A**

**x**

**x\_avg**

**x\_dev**

**y**

**z**

**x\_cen**

**x\_avg\_cen**

**y\_std**

**z\_std**

**C:\Howtodo\DataS1.txt !SKIP 1**

**!MVREMOVE**

**!MAXIT 5000**

**!SLOW 3**

**!EXTRA 5**

**!CONTINUE**

**y z ~ Trait !r Trait.Indiv**

**1 2 1 !STEP 0.001**

**0**

**Trait 0 US !GPUP !S2==1**

**0.7454**

**0.000 0.7764**

**Trait.Indiv 2**

**Trait 0 US !GPUP**

**0.3215**

**0.000 0.2575**

**Indiv**

Part 2: coding of the .pin file

**# Pin file for bivariate MM with Individual as random effect**

**# (Co)variance components given in output are**

**# 1. trait A residual error variance**

**# 2. residual covariance A & B**

**# 3. trait B residual error variance**

**# 4. trait A individual variance**

**# 5. individual covariance A & B**

**# 6. trait B individual variance**

**F PhenVarA 1+4 # 7. Phenotypic variance of A with SE**

**F PhenVarB 3+6 # 8. Phenotypic variance of B with SE**

**F ResVarA 1 # 9. Residual variance of A with SE**

**F ResVarB 3 # 10. Residual variance of B with SE**

**F IndVarA 4 # 11. Individual variance of A with SE**

**F IndVarB 6 # 12. Individual variance of B with SE**

**H RepeatA 4 7 # Repeatability of A with SE**

**H RepeatB 6 8 # Repeatability of B with SE**

**F PhenCovAB 2+5 # 13. Phenotypic covariance AB with SE**

**F ResCovAB 2 # 14. Residual covariance AB with SE**

**F IndCovAB 5 # 15. Individual covariance AB with SE**

**R PhenCorAB 7 13 8 # Phenotypic correlation AB with SE**

**R ResCorAB 1 2 3 # Residual correlation AB with SE**

**R IndCorAB 4 5 6 # Individual correlation AB with SE**

Part 3: Key output reported in the .asr file

**36 LogL=-434.832 S2= 1.0000 498 df**

**- - - Results from analysis of y z - - -**

**Source Model terms Gamma Component Comp/SE % C**

**Residual UnStructured 1 1 1.70446 1.70446 10.00 0 P**

**Residual UnStructured 2 1 -0.826967E-02 -0.826967E-02 -0.07 0 U**

**Residual UnStructured 2 2 1.51714 1.51714 10.00 0 P**

**Trait.Indiv UnStructured 1 1 1.03295 1.03295 3.69 0 P**

**Trait.Indiv UnStructured 2 1 0.639332 0.639332 3.06 0 U**

**Trait.Indiv UnStructured 2 2 0.939411 0.939411 3.71 0 P**

**Covariance/Variance/Correlation Matrix UnStructured Residual**

**1.704 -0.5143E-02**

**-0.8270E-02 1.517**

**Covariance/Variance/Correlation Matrix UnStructured Trait.Indiv**

**1.033 0.6490**

**0.6393 0.9394**

**Wald F statistics**

**Source of Variation NumDF DenDF F-inc P-inc**

**11 Trait 2 48.0 57.89 <.001**

**Solution Standard Error T-value T-prev**

**11 Trait**

**1 -0.590446 0.165761 -3.56**

**2 -1.68751 0.157660 -10.70 -6.70**

**12 Trait.Indiv 100 effects fitted**

**Finished: 02 Mar 2011 08:27:53.723 LogL Converged**

Part 4: Variance components reported in the .pvc file

**7 PhenVarA 1 2.7374 0.30924**

**8 PhenVarB 3 2.4566 0.27889**

**9 ResVarA 1 1.7045 0.17045**

**10 ResVarB 3 1.5171 0.15171**

**11 IndVarA 4 1.0330 0.27964**

**12 IndVarB 6 0.93941 0.25292**

**RepeatA = Trait.In 4/PhenVarA 7= 0.3773 0.0704**

**RepeatB = Trait.In 6/PhenVarB 8= 0.3824 0.0704**

**13 PhenCovAB 2 0.63106 0.22676**

**14 ResCovAB 2 -0.82697E-02 0.11371**

**15 IndCovAB 5 0.63933 0.20895**

**PhenCorAB = PhenCovA/SQR[PhenVarA\*PhenVarB]= 0.2434 0.0751**

**ResCorAB = Residual/SQR[Residual\*Residual]= -0.0051 0.0707**

**IndCorAB = Trait.In/SQR[Trait.In\*Trait.In]= 0.6490 0.1359**

**Notice: The parameter estimates are followed by**

**their approximate standard errors.**

Part 5: Overview of statistical results (and where to find them)

**±SE = -0.590446±0.165761 (asr file)**

**±SE = -1.68751±0.157660 (asr file)**

**±SE = 2.7374±0.30924 (pvc file)**

**±SE = 2.4566±0.27889 (pvc file)**

**±SE =1.0330±0.27964 (pvc file)**

**±SE = 0.93941±0.25292 (pvc file)**

**±SE = 1.7045±0.17045 (pvc file)**

**±SE = 1.5171±0.15171 (pvc file)**

**Repeatabilityy±SE = 0.3773±0.0704 (pvc file)**

**Repeatabilityz±SE = 0.3824±0.0704 (pvc file)**

**±SE = 0.63106±0.22676 (pvc file)**

**±SE = 0.63933±0.20895 (pvc file)\***

**±SE = -0.82697E-02±0.11371 (pvc file)\***

**±SE = 0.2434±0.0751 (pvc file)**

**±SE = 0.6490±0.1359 (pvc file)\***

**±SE =-0.0051±0.0707 (pvc file)\***

**Model Loglikelihood = -434.832 (asr file).**

**\*see section E how to calculate significance using a likelihood ratio test**

*R coding, output, and interpretation (4 parts)*

Part 1: coding for MCMCglmm(Hadfield 2010) R library

#MCMC fitted model using the MCMCglmm library

#using inverse gamma prior and longer than default

#iterations. Please refer to Hadfield (2010) and the

#primary literature on priors for further discussion of

#priors.

prior.bivar<-list(R=list(V=diag(2),nu=1.002),G=list(G1=list(V=diag(2), nu=1.002)))

yz.bivarC.mcmc<-

MCMCglmm(cbind(y,z)~(trait-1),

random=~us(trait):indiv,rcov=~us(trait):units,

family=c("gaussian","gaussian"), prior=prior.bivar,

nitt=1300000,thin=1000,burnin=300000, data=set1.data,verbose=FALSE)

Part 2: key output

from summary(yz.bivarC.mcmc), omits fixed effects information:

DIC: 1733.672

G-structure: ~us(trait):indiv

post.mean l-95% CI u-95% CI eff.samp

y:y.indiv 1.0838 0.5864 1.660 1000.0

z:y.indiv 0.6443 0.2526 1.057 1000.0

y:z.indiv 0.6443 0.2526 1.057 1000.0

z:z.indiv 1.0010 0.5558 1.527 991.8

R-structure: ~us(trait):units

post.mean l-95% CI u-95% CI eff.samp

y:y.units 1.73560 1.3958 2.0695 1000

z:y.units -0.01233 -0.2596 0.1967 1000

y:z.units -0.01233 -0.2596 0.1967 1000

z:z.units 1.54341 1.2150 1.8336 1111

Part 3: Calculation of repeatabilities

MCMCglmm coding:

rep.y<-

yz.bivarC.mcmc$VCV[,1]/ (yz.bivarC.mcmc$VCV[,1]+ yz.bivarC.mcmc$VCV[,5])

rep.z<-

yz.bivarC.mcmc$VCV[,4]/ (yz.bivarC.mcmc$VCV[,4]+ yz.bivarC.mcmc$VCV[,8])

#posterior mode for repeatability of each response:

posterior.mode(rep.y)

posterior.mode(rep.z)

#95% credibility intervals for repeatabilities:

HPDinterval(rep.y)

HPDinterval(rep.z)

Part 4: Overview of parameter estimates (and how to retrieve them)

posterior modes between credibility intervals:

: -0.926 < -0.554 < -0.244 (posterior.mode(yz.bivarC.mcmc$Sol;

HPDinterval(yz.bivarC.mcmc$Sol))

: -2.007 < -1.695 < -1.380 (same as for )

: 0.586 < 1.198 < 1.660 (posterior.mode(yz.bivarC.mcmc$VCV); HPDinterval(yz.bivarC.mcmc$VCV))

: 0.556 < 0.888 < 1.527 (same as for )

: 1.396 < 1.749 < 2.069 (same as for )

: 1.215 < 1.525 < 1.834 (same as for )

Repeatabilityy: 0.259 < 0.365 < 0.518 (see Part 3)

Repeatabilityz: 0.263 < 0.366 < 0.524 (see Part 3)

: 0.252 < 0.598 < 1.057 (same as for )

: -0.260 < -0.001 < 0.197 (same as for )

: = 0.579 (can also use the posterior.cor function)

: = -0.001 (can also use the posterior.cor function)

*SAS 9.2 coding, output, and interpretation* *(4 parts)*

Part 1: Background

PROC MIXED in SAS 9.2 is not directly designed to do multivariate models. However, it is flexible enough to produce the proper analysis. However, the structure of input data must be adjusted to allow bivariate modeling. Instead of both traits being separate columns of data, we stacked the data in one column and created a separate column to indicate which trait a value is. Thus, we integrated y and z in the same column (called “RESPONSE”) and created an indicator variable (called “TRAIT”) that indicates which trait (Y or Z) is in RESPONSE. In addition, a new variable needs to be created that links observations taken at the same time, which in this file ranges between 1 and 5. We called this “OBS”. These manipulations can be done in a data spreadsheet package or in SAS itself. In this case a new spreadsheet was created called “DataS1tall” and uploaded into SAS.

Part 2: Model statement

**Proc** **mixed** data=HowTO.DataS1tall Method=reml covtest;

class INDIV TRAIT OBS;

model Response=TRAIT /noint solution;

random TRAIT /type=un sub=INDIV g;

repeated TRAIT OBS /type=un@cs sub=INDIV r;

**run**;

This code contains key instructions for generating the within-individual covariance matrix and instructions for presenting the results. In the random statement, the “g” at the end of the line instructs SAS to output the G matrix which will contain the between-individual variances for each trait and the between-individual between-trait covariance. The repeated statement instructs SAS to generate a nested R matrix, in which within-individual, within-trait variance is calculated as well as the within-individual between-trait covariance. This is accomplished by the double repeated statement involving TRAIT and OBS and the “TYPE = UN@CS” command. The R matrix is outputted by adding “r” at the end of the line.

Part 3: Output

Estimated R Matrix

Row Col1 Col2 Col3 Col4 Col5 Col6 Col7 Col8 Col9 Col10

1 1.7045 -0.00827

2 1.7045 -0.00827

3 1.7045 -0.00827

4 1.7045 -0.00827

5 1.7045 -0.00827

6 -0.00827 1.5171

7 -0.00827 1.5171

8 -0.00827 1.5171

9 -0.00827 1.5171

10 -0.00827 1.5171

Estimated G Matrix

Row Effect Trait Indiv Col1 Col2

1 Trait Y 1 1.0330 0.6393

2 Trait Z 1 0.6393 0.9394

Covariance Parameter Estimates

Standard Z

Cov Parm Subject Estimate Error Value Pr Z

UN(1,1) Indiv 1.0330 0.2796 3.69 0.0001

UN(2,1) Indiv 0.6393 0.2090 3.06 0.0022

UN(2,2) Indiv 0.9394 0.2529 3.71 0.0001

Trait UN(1,1) Indiv 1.7045 0.1704 10.00 <.0001

UN(2,1) Indiv -0.00827 0.1137 -0.07 0.9420

UN(2,2) Indiv 1.5171 0.1517 10.00 <.0001

Obs Corr Indiv 0 . . .

Fit Statistics

-2 Res Log Likelihood 1784.9

AIC (smaller is better) 1798.9

AICC (smaller is better) 1799.2

BIC (smaller is better) 1812.3

Solution for Fixed Effects

Standard

Effect Trait Estimate Error DF t Value Pr > |t|

Trait Y -0.5904 0.1658 98 -3.56 0.0006

Trait Z -1.6875 0.1577 98 -10.70 <.0001

Part 4: Overview of statistical results (and where to find them)

±SE = -0.5904±0.1658 (“Trait Y” from table “Solution for Fixed Effects)

±SE = -1.6875±0.1577 (“Trait Z” from table “Solution for Fixed Effects)

±SE = 2.7375 (“UN(1,1)” + “Trait UN(1,1)” in table “Covariance Parameter Estimates”) [We hope to provide information on the calculation of the uncertainty (standard errors) around in updated versions of this text on our website.]

±SE = 2.4550 (“UN(2,2)” + “Trait UN(2,2)” in table “Covariance Parameter Estimates”)

[We hope to provide information on the calculation of the uncertainty (standard errors) around in updated versions of this text on our website.]

±SE =1.0330±0.2796 (“UN(1,1)” in table “Covariance Parameter Estimates”)

±SE = 0.9394±0.2529 (“UN(2,2)” in table “Covariance Parameter Estimates”)

±SE = 1.7045±0.1704 (“Trait UN(1,1)” in table “Covariance Parameter Estimates”)

±SE = 1.5171±0.1517 (“Trait UN(2,2)” in table “Covariance Parameter Estimates”)

±SE = 0.6377±0.2077 (Not shown above; obtained by running model in Part C below with these data, in which case entry “UN(2,1) in table “Covariance Parameter Estimates” becomes the phenotypic covariance)

±SE = 0.63933±0.20895\* (“UN(2,1)” in table “Covariance Parameter Estimates”)

±SE = -0.82697E-02±0.11371\* (“Trait UN(2,1)” in table “Covariance Parameter Estimates”)

Model Log likelihood = -892.4 (“-2res Log Likelihood” divided by -2, from table of fit statistics)\*

The following parameters must be calculated from the above either through extra SAS programming or in a spreadsheet program such as excel using the appropriate formulas:

Repeatabilityy, Repeatabilityz may be calculated by hand by filling in the estimated variance components in Eqn. 2; we hope to provide information on the calculation of repeatability and its uncertainty (standard errors) in updated versions of this text on our website.

, \* and \* may be calculated by hand by filling in the estimated variance components in Eqn. 7, Eqn. 7c, and Eqn. 7d respectively; we hope to provide information on the calculation of those correlations and their uncertainty (standard errors) in updated versions of this text on our website.

\*see section E how to calculate significance using a likelihood ratio test

C. BIVARIATE MMs FOR A SCENARIO WHERE TWO PHENOTYPIC ATTRIBUTES WERE BOTH ASSAYED REPEATEDLY BUT NEVER AT THE SAME TIME (SCENARIO 4; TABLE 2)

*Application*: The example is for running a bivariate MM with random intercepts for the factor “Individual”. , , , and calculating repeatability *y* and *z*, with their associated confidence (standard errors / credible intervals). The phenotypic attributes *y* and *z* are never assayed at the same time: / can be estimated but not / ) The example uses *DataS2.txt*; the phenotypic attributes *y* and *z* are used as the response variables.

*Specified model: Linear Eqn. 7a and random effect structure similar to Eqn. 7b but with constrained to zero:*

:

:

*ASReml coding, output, and interpretation (5 parts)*

Part 1: coding of the .as file

**ASREML analysis**

**Indiv !A**

**y**

**z**

**y\_std**

**z\_std**

**C:\Howtodo\DataS2.txt !SKIP 1**

**!MVREMOVE**

**!MAXIT 5000**

**!SLOW 3**

**!EXTRA 5**

**!CONTINUE**

**y z ~ Trait !r Trait.Indiv**

**1 2 1 !STEP 0.001**

**0**

**Trait 0 US !GPZP !S2==1**

**0.7454**

**0.000 0.7764**

**Trait.Indiv 2**

**Trait 0 US !GPUP**

**0.3215**

**0.000 0.2575**

**Indiv**

Part 2: coding of the .pin file

**# Pin file for bivariate MM with Individual as random effect and with non-estimable residual covariance**

**# (Co)variance components given in output are**

**# 1. trait A residual error variance**

**# 2. trait B residual error variance**

**# 3. trait A individual variance**

**# 4. individual covariance A & B**

**# 5. trait B individual variance**

**F PhenVarA 1+3 # 6. Phenotypic variance of A with SE**

**F PhenVarB 2+5 # 7. Phenotypic variance of B with SE**

**F ResVarA 1 # 8. Residual variance of A with SE**

**F ResVarB 2 # 9. Residual variance of B with SE**

**F IndVarA 3 # 10. Individual variance of A with SE**

**F IndVarB 5 # 11. Individual variance of B with SE**

**H RepeatA 3 6 # Repeatability of A with SE**

**H RepeatB 5 7 # Repeatability of B with SE**

**F IndCovAB 4 # 12. Individual covariance AB with SE**

**R IndCorAB 3 4 5 # Individual correlation AB with SE**

Part 3: Key output reported in the .asr file

**36 LogL=-434.835 S2= 1.0000 498 df**

**- - - Results from analysis of y z - - -**

**Source Model terms Gamma Component Comp/SE % C**

**Residual UnStructured 1 1 1.70446 1.70446 10.00 0 P**

**Residual UnStructured 2 2 1.51714 1.51714 10.00 0 P**

**Trait.Indiv UnStructured 1 1 1.03295 1.03295 3.69 0 P**

**Trait.Indiv UnStructured 2 1 0.637678 0.637678 3.07 0 U**

**Trait.Indiv UnStructured 2 2 0.939411 0.939411 3.71 0 P**

**Covariance/Variance/Correlation Matrix UnStructured Residual**

**1.704 0.000**

**0.000 1.517**

**Covariance/Variance/Correlation Matrix UnStructured Trait.Indiv**

**1.033 0.6473**

**0.6377 0.9394**

**Wald F statistics**

**Source of Variation NumDF DenDF F-inc P-inc**

**6 Trait 2 48.0 57.89 <.001**

**Notice: The DenDF values are calculated ignoring fixed/boundary/singular**

**variance parameters using numerical derivatives.**

**Solution Standard Error T-value T-prev**

**6 Trait**

**1 -0.590446 0.165761 -3.56**

**2 -1.68751 0.157660 -10.70 -6.70**

**7 Trait.Indiv 100 effects fitted**

**Finished: 02 Mar 2011 09:16:50.280 LogL Converged**

Part 4: Variance components reported in the .pvc file

**6 PhenVarA 1 2.7374 0.30924**

**7 PhenVarB 2 2.4566 0.27889**

**8 ResVarA 1 1.7045 0.17045**

**9 ResVarB 2 1.5171 0.15171**

**10 IndVarA 3 1.0330 0.27964**

**11 IndVarB 5 0.93941 0.25292**

**RepeatA = Trait.In 3/PhenVarA 6= 0.3773 0.0704**

**RepeatB = Trait.In 5/PhenVarB 7= 0.3824 0.0704**

**12 IndCovAB 4 0.63768 0.20771**

**IndCorAB = Trait.In/SQR[Trait.In\*Trait.In]= 0.6473 0.1339**

**Notice: The parameter estimates are followed by**

**their approximate standard errors.**

Part 5: Overview of statistical results (and where to find them)

**±SE = -0.590446±0.165761 (asr file)**

**±SE = -1.68751±0.157660 (asr file)**

**±SE = 2.7374±0.30924 (pvc file)**

**±SE = 2.4566±0.27889 (pvc file)**

**±SE =1.0330±0.27964 (pvc file)**

**±SE = 0.93941±0.25292 (pvc file)**

**±SE = 1.7045±0.17045 (pvc file)**

**±SE = 1.5171±0.15171 (pvc file)**

**Repeatabilityy±SE = 0.3773±0.0704 (pvc file)**

**Repeatabilityz±SE = 0.3824±0.0704 (pvc file)**

**±SE = 0.63768±0.20771 (pvc file)\***

**±SE = 0.6473±0.1339 (pvc file)\***

**Model Loglikelihood = -434.835 (asr file).**

**\*see section E how to calculate significance using a likelihood ratio test**

*R coding, output, and interpretation (4 parts)*

Part 1: coding for MCMCglmm(Hadfield 2010) R library

#MCMC fitted model using the MCMCglmm library

#using an improper prior and longer than default

#iterations. Please refer to Hadfield (2010) and the

#primary literature on priors for further discussion of

#priors. Note that the rcov specification (the within-

#individual specification) now uses “idh()” to specify

#the matrix format. This is equivalent to SAS’s banded

#main diagonal matrix and forces the within-individual

#covariances to be zero.

prior.bivar<-list(R=list(V=diag(2),nu=1.002),G=list(G1=list(V=diag(2), nu=1.002)))

yz.bivarD.mcmc<-MCMCglmm(cbind(y,z)~(trait-1),

random=~us(trait):indiv,rcov=~idh(trait):units,

family=c("gaussian","gaussian"), prior=prior.bivar,

nitt=130000,thin=100,burnin=30000, data=set2.data,verbose=FALSE)

Part 2: key output

from summary(yz.bivarD.mcmc), omits fixed effects information:

DIC: 1731.828

G-structure: ~us(trait):indiv

post.mean l-95% CI u-95% CI eff.samp

y:y.indiv 1.0932 0.5404 1.647 1000

z:y.indiv 0.6387 0.2145 1.073 1490

y:z.indiv 0.6387 0.2145 1.073 1490

z:z.indiv 1.0029 0.5412 1.602 1306

R-structure: ~idh(trait):units

post.mean l-95% CI u-95% CI eff.samp

y.units 1.720 1.383 2.051 1123

z.units 1.532 1.254 1.861 1103

Part 3: Calculation of repeatabilities

MCMCglmm coding:

rep.y<-

yz.bivarD.mcmc$VCV[,1]/ (yz.bivarD.mcmc$VCV[,1]+ yz.bivarD.mcmc$VCV[,5])

rep.z<-

yz.bivarD.mcmc$VCV[,4]/ (yz.bivarD.mcmc$VCV[,6]+ yz.bivarD.mcmc$VCV[,6])

#posterior mode for repeatability of each response:

posterior.mode(rep.y)

posterior.mode(rep.z)

#95% credibility intervals for repeatabilities:

HPDinterval(rep.y)

HPDinterval(rep.z)

Part 4: Overview of parameter estimates (and how to retrieve them)

posterior modes between credibility intervals:

: -0.912 < -0.567 < -0.251 (posterior.mode(yz.bivarD.mcmc$Sol;

HPDinterval(yz.bivarD.mcmc$Sol))

: -1.996 < -1.682 < -1.386 (same as for )

: 0.540 < 0.999 < 1.647 (posterior.mode(yz.bivarD.mcmc$VCV); HPDinterval(yz.bivarD.mcmc$VCV))

: 0.541 < 0.872 < 1.602 (same as for )

: 1.383 < 1.769 < 2.051 (same as for )

: 1.254 < 1.504 < 1.861 (same as for )

Repeatabilityy: 0.249 < 0.347 < 0.515 (see Part 3)

Repeatabilityz: 0.259 < 0.368 < 0.541 (see Part 3)

: 0.215 < 0.649 < 1.073 (same as for )

: = 0.695 (can also use the posterior.cor function)

*SAS 9.2 coding, output, and interpretation* *(4 parts)*

Part 1: Background

This procedure requires that Dataset S2 be modified by again stacking the data for each trait in one column, calling that column “RESPONSE”, and then coding a second column called “TRAIT” indicating whether the measurement is from trait y or z. Because no within-individual, between trait covariance can be calculated, there is no need to add a variable coding observation as in Part B. The new dataset is called “DataS2tall”.

Part 2: Code

**Proc** **mixed** data=HOWTO.DataS2tall method=reml covtest;

class INDIV TRAIT;

model RESPONSE=TRAIT/solution;

random TRAIT /type=un sub=INDIV g;

repeated /type=vc grp=TRAIT sub=INDIV;

**run**;

Part 3: Output

Estimated G Matrix

Row Effect Trait Indiv Col1 Col2

1 Trait y 1 1.0330 0.6377

2 Trait z 1 0.6377 0.9394

Covariance Parameter Estimates

Standard Z

Cov Parm Subject Group Estimate Error Value Pr Z

UN(1,1) Indiv 1.0330 0.2796 3.69 0.0001

UN(2,1) Indiv 0.6377 0.2077 3.07 0.0021

UN(2,2) Indiv 0.9394 0.2529 3.71 0.0001

Residual Indiv Trait y 1.7045 0.1704 10.00 <.0001

Residual Indiv Trait z 1.5171 0.1517 10.00 <.0001

Fit Statistics

-2 Res Log Likelihood 1784.9

AIC (smaller is better) 1794.9

AICC (smaller is better) 1795.1

BIC (smaller is better) 1804.5

Solution for Fixed Effects

Standard

Effect Trait Estimate Error DF t Value Pr > |t|

Trait y -0.5904 0.1658 98 -3.56 0.0006

Trait z -1.6875 0.1577 98 -10.70 <.0001

Part 4: Overview of statistical results (and where to find them)

±SE = -0.5904±0.1658 (“Trait Y” from table “Solution for Fixed Effects)

±SE = -1.6875±0.1577 (“Trait Z” from table “Solution for Fixed Effects)

±SE = 2.7375 (“UN(1,1)” + “Trait UN(1,1)” in table “Covariance Parameter Estimates”)

[We hope to provide information on the calculation of the uncertainty (standard errors) around in updated versions of this text on our website.]

±SE = 2.4550 (“UN(2,2)” + “Trait UN(2,2)” in table “Covariance Parameter Estimates”) [We hope to provide information on the calculation of the uncertainty (standard errors) around in updated versions of this text on our website.]

±SE =1.0330±0.2796 (“UN(1,1)” in table “Covariance Parameter Estimates”)

±SE = 0.9394±0.2529 (“UN(2,2)” in table “Covariance Parameter Estimates”)

±SE = 1.7045±0.1704 (“Residual Indiv Trait Y” in table “Covariance Parameter Estimates”)

±SE = 1.5171±0.1517 (“Residual Indiv Trait Z” in table “Covariance Parameter Estimates”)

±SE = 0.6377±0.2077 (“UN(2,1)” in table “Covariance Parameter Estimates”)

±SE = (calculated using Eqn. 7c and parameters above)\*

Model Log likelihood = -892.4 (“-2res Log Likelihood” divided by -2, from table of fit statistics)\*

\*see section E how to calculate significance using a likelihood ratio test

D. BIVARIATE MMs TO ESTIMATE REPEATABILITY FOR TWO DATASETS SIMULTANEOUSLY

*Application*: The example is for running a bivariate MM with random intercepts for the factor “Individual”, where *y* and *z* represent the same phenotypic attribute measured for different classes of individuals (e.g. females and males, respectively). , , , and repeatability are calculated for *y* and *z*, with their associated confidence (standard errors / credible intervals). In the example, individuals cannot change sex; neither / nor / can therefore be estimated. The example uses *DataS3.txt*. We expressed both *y* and *z* in standard deviation units (i.e. and ), such that values of and thus represent the repeatability of *y* and *z*, respectively; *y\_std* and *z\_std* are the response variables. We use here standardized response variables because by doing so we can use the output to subsequently address the question of whether repeatability differs between *y* and *z* (detailed in section FE).

*Specified model: Linear Eqn. 8a and random effect structure Eqn. S1:*

:

:

*ASReml coding, output, and interpretation (5 parts)*

Part 1: coding of the .as file

**ASREML analysis**

**Indiv !A**

**y**

**z**

**y\_std**

**z\_std**

**C:\Howtodo\DataS3.txt !SKIP 1**

**!MVREMOVE**

**!MAXIT 5000**

**!SLOW 3**

**!EXTRA 5**

**!CONTINUE**

**y\_std z\_std ~ Trait !r Trait.Indiv**

**1 2 1 !STEP 0.001**

**0**

**Trait 0 US !GPZP !S2==1**

**0.7454**

**0.000 0.7764**

**Trait.Indiv 2**

**Trait 0 US !GPZP**

**0.3215**

**0.000 0.2575**

**Indiv**

Part 2: coding of the .pin file

**# Pin file for bivariate MM with Individual as random effect with all covariances non-estimable**

**# (Co)variance components given in output are**

**# 1. trait A residual error variance**

**# 2. trait B residual error variance**

**# 3. trait A individual variance**

**# 4. individual covariance A & B (printed though it is non-estimable)**

**# 5. trait B individual variance**

**F PhenVarA 1+3 # 6. Phenotypic variance of A with SE**

**F PhenVarB 2+5 # 7. Phenotypic variance of B with SE**

**F ResVarA 1 # 8. Residual variance of A with SE**

**F ResVarB 2 # 9. Residual variance of B with SE**

**F IndVarA 3 # 10. Individual variance of A with SE**

**F IndVarB 5 # 11. Individual variance of B with SE**

**H RepeatA 3 6 # Repeatability of A with SE**

**H RepeatB 5 7 # Repeatability of B with SE**

Part 3: Key output reported in the .asr file

**36 LogL=-205.755 S2= 1.0000 498 df**

**- - - Results from analysis of y\_std z\_std - - -**

**Source Model terms Gamma Component Comp/SE % C**

**Residual UnStructured 1 1 0.626451 0.626451 10.00 0 P**

**Residual UnStructured 2 2 0.621406 0.621406 10.00 0 P**

**Trait.Indiv UnStructured 1 1 0.379647 0.379647 3.69 0 P**

**Trait.Indiv UnStructured 2 1 0.00000 0.00000 0.00 0 F**

**Trait.Indiv UnStructured 2 2 0.384775 0.384775 3.71 0 P**

**Covariance/Variance/Correlation Matrix UnStructured Residual**

**0.6265 0.000**

**0.000 0.6214**

**Covariance/Variance/Correlation Matrix UnStructured Trait.Indiv**

**0.3796 0.000**

**0.000 0.3848**

**Wald F statistics**

**Source of Variation NumDF DenDF F-inc P-inc**

**6 Trait 2 64.6 0.00 1.000**

**Solution Standard Error T-value T-prev**

**6 Trait**

**1 0.00000 0.100493 0.00**

**2 0.00000 0.100902 0.00 0.00**

**7 Trait.Indiv 200 effects fitted ( 100 are zero)**

**Finished: 02 Mar 2011 09:26:37.492 LogL Converged**

Part 4: Variance components reported in the .pvc file

**6 PhenVarA 1 1.0061 0.11366**

**7 PhenVarB 2 1.0062 0.11423**

**8 ResVarA 1 0.62645 0.62645E-01**

**9 ResVarB 2 0.62141 0.62141E-01**

**10 IndVarA 3 0.37965 0.10278**

**11 IndVarB 5 0.38477 0.10359**

**RepeatA = Trait.In 3/PhenVarA 6= 0.3773 0.0704**

**RepeatB = Trait.In 5/PhenVarB 7= 0.3824 0.0704**

**Notice: The parameter estimates are followed by**

**their approximate standard errors.**

Part 5: Overview of statistical results

**±SE = 0.00000±0.100493 (asr file)**

**±SE = 0.00000±0.100902 (asr file)**

**±SE = 1.0061±0.11366 (pvc file)**

**±SE = 1.0062±0.11423 (pvc file)**

**±SE = 0.37965±0.10278 (pvc file)**

**±SE = 0.38477±0.10359 (pvc file)**

**±SE = 0.62645±0.62645E-01 (pvc file)**

**±SE = 0.62141±0.62141E-01 (pvc file)**

**Repeatabilityy±SE = 0.3773±0.0704 (pvc file)\***

**Repeatabilityz±SE = 0.3824±0.0704 (pvc file)\***

**Model Loglikelihood = -205.755 (asr file).**

**\*see section F how to calculate significance of the difference in repeatability between *y* and *z* using a likelihood ratio test**

*R coding, output, and interpretation (4 parts)*

Part 1: coding for MCMCglmm(Hadfield 2010) R library

#Please refer to Hadfield (2010) and the

#primary literature on priors. Note that the rcov

#and random specification now both use “idh()” to specify

#the matrix format. This is equivalent to SAS’s banded

#main diagonal matrix and forces the

#covariances to be zero.

yz.bivarE.mcmc<-MCMCglmm(cbind(y\_std,z\_std)~(trait-1),

random=~idh(trait):indiv,rcov=~idh(trait):units,

family=c("gaussian","gaussian"), prior=prior.bivar,

nitt=130000,thin=100,burnin=30000, data=set3.data,verbose=FALSE)

Part 2: key output

from summary(yz.bivarE.mcmc), omits fixed effects information:

DIC: 1264.73

G-structure: ~idh(trait):indiv

post.mean l-95% CI u-95% CI eff.samp

y\_std.indiv 0.4137 0.2051 0.6376 1000

z\_std.indiv 0.4230 0.2367 0.6645 1256

R-structure: ~idh(trait):units

post.mean l-95% CI u-95% CI eff.samp

y\_std.units 0.6329 0.5162 0.7496 1000

z\_std.units 0.6324 0.5206 0.7663 1000

Part 3: Calculation of repeatabilities

MCMCglmm coding:

rep.y<-

yz.bivarE.mcmc$VCV[,1]/ (yz.bivarE.mcmc$VCV[,1]+ yz.bivarE.mcmc$VCV[,3])

rep.z<-

yz.bivarE.mcmc$VCV[,2]/ (yz.bivarE.mcmc$VCV[,2]+ yz.bivarE.mcmc$VCV[,4])

#posterior mode for repeatability of each response:

posterior.mode(rep.y)

posterior.mode(rep.z)

#95% credibility intervals for repeatabilities:

HPDinterval(rep.y)

HPDinterval(rep.z)

Part 4: Overview of parameter estimates (and how to retrieve them)

posterior modes between credibility intervals:

: -0.205 < 0.026 < 0.218 (posterior.mode(yz.bivarE.mcmc$Sol;

HPDinterval(yz.bivarE.mcmc$Sol))

: -0.212 < -0.019 < 0.210 (same as for )

: 0.205 < 0.450 < 0.638 (posterior.mode(yz.bivarE.mcmc$VCV); HPDinterval(yz.bivarE.mcmc$VCV))

: 0.237 < 0.380 < 0.664 (same as for )

: 0.516 < 0.620 < 0.750 (same as for )

: 0.524 < 0.626 < 0.766 (same as for )

Repeatabilityy: 0.262 < 0.406 < 0.525 (see Part 3)

Repeatabilityz: 0.266 < 0.388 < 0.533 (see Part 3)

*SAS 9.2 coding, output, and interpretation* *(4 parts)*

Part 1: Explanation

There are two ways in SAS to calculate repeatabilities of separate datasets. Below we give the bivariate method in parallel to the other two programs above. The second method is presented in Section E in the likelihood ratio test of the hypothesis that the two repeatabilities are the same. We note that either procedure can be used to test repeatabilities of two different traits or of the same trait in two different groups.

Part 2: Code

**Proc** **mixed** data=HowTO.DataS3tall Method=reml cl alpha=**.05** covtest;

class INDIV TRAIT;

model STDRESP=TRAIT/noint;

random TRAIT / Type = un sub=INDIV g;

repeated /type=vc grp=TRAIT sub=INDIV;

**run**;

Part 3: output

Estimated G Matrix

Row Effect Trait Indiv Col1 Col2

1 Trait Y 1 0.3796

2 Trait Z 1 0.3848

Covariance Parameter Estimates

Standard Z

Cov Parm Subject Group Estimate Error Value Pr Z Alpha Lower Upper

UN(1,1) Indiv 0.3796 0.1028 3.69 0.0001 0.05 0.2378 0.7007

UN(2,1) Indiv 0 . . . . . .

UN(2,2) Indiv 0.3848 0.1036 3.71 0.0001 0.05 0.2416 0.7074

Residual Indiv Trait Y 0.6265 0.06265 10.00 <.0001 0.05 0.5198 0.7699

Residual Indiv Trait Z 0.6214 0.06214 10.00 <.0001 0.05 0.5156 0.7637

Fit Statistics

-2 Res Log Likelihood 1326.8

AIC (smaller is better) 1336.8

AICC (smaller is better) 1336.9

BIC (smaller is better) 1349.8

Part 4: Analysis and Interpretation

Because standardized trait values were used in this analysis, the between individual variances are repeatabilities, and the two values are listed in the first and third line of the table “Covaraince Parameter Estimates”. Because they were measured in different individuals, there is no between-individual covariance (i.e., “UN(2,1)” = 0).

E. LIKELIHOOD RATIO TESTS

*Application*: The statistical significance of variances, repeatabilities, and covariances (correlations) can be evaluated by using a likelihood ratio test (LRT) (Meyer 1992; Wilson *et al.* 2010). LRTs may also be used to calculate the significance of differences between datasets in (co)variance components or repeatabilities. Here we apply LRTs for estimating the significances of (co)variances, and differences in repeatability across datasets, for the examples discussed in sections A – D.

LRTs are calculated using a three-step approach. First, the unconstrained model is fitted (see sections A – D above). Second, an alternative model is fitted where specific (co)variance components are constrained. In this section we provide the associated code for such constrained models (see sections EA – ED below) such that the LRT can be calculated for the examples discussed in sections A – D. Third, the *χ2*-distributed LRT is calculated as two times the absolute difference in LogLikelihood between the constrained and unconstrained model; the number of degrees of freedom equals the difference in the number of estimated parameters between the two models (Meyer 1992).

Alternatively, statistical support for a repeatability, variance, or covariance can be determined by differences in values of information criteria (Grueber *et al.* 2011). We demonstrate this approach for *R* using output from MCMCglmm model fitting. Specifically we use the deviance information criteria (DIC) to determine the support for a model including or excluding a particular term. While “statistical significance” is not determined with DIC values, the models with lower DIC values have a better fit to the data when penalized for complexity.

EA. LRT-BASED SIGNIFICANCE OF IN UNIVARIATE MMs

*Goal*: LRT-based significance calculation of for the model described in section “A. UNIVARIATE MMs WITH RANDOM INTERCEPTS FOR “INDIVIDUAL”.

*ASReml coding, and calculation of significance (2 parts)*

Part 1: coding of the constrained .as file

**ASREML analysis**

**Indiv !A**

**x**

**x\_avg**

**x\_dev**

**y**

**z**

**x\_cen**

**x\_avg\_cen**

**y\_std**

**z\_std**

**C:\Howtodo\DataS1.txt !SKIP 1**

**!MVREMOVE**

**!MAXIT 5000**

**!SLOW 3**

**!EXTRA 5**

**!CONTINUE**

**y ~ mu x\_avg\_cen x\_dev**

**1 1 0 !STEP 0.001**

**0 0 ID 0.1 !S2==1**

Part 2: Calculation of significance of

1. **Loglikelihood Unconstrained = -229.472 (asr file of unconstrained model)**
2. **Loglikelihood Constrained = -253.565 (asr file of constrained model – not printed)**
3. **LRT: *χ21=48.186, P < 0.001***

*R coding, output, and interpretation (2 parts)*

Part 1: coding of the constrained model

y.alt.mcmc<-MCMCglmm(y~x\_avg\_cen+x\_dev,

family="gaussian",

data=set1.data, verbose=FALSE)

Part 2: Determining statistical support for or against the unconstrained model

1. DIC Unconstrained: 885.48 (From A)
2. DIC Constrained: 965.125 (from Part 1, this section)
3. DIC Unconstrained – DIC Constrained: -79.945

Based on the difference (-79.945), the unconstrained model is considerably more well supported by the data.

*SAS coding, output, and interpretation (4 parts): ML-method*

Part 1: Background

Extract by hand the -2 log likelihood from the fit statistics table in the output of two models (the model in Part A above must be modified to be maximum likelihood instead of restricted ML, and is labelled model 2. Model 1 is presented below). Look up difference in -2LL between the two models in a chi-square table (or web site, http://faculty.vassar.edu/lowry/tabs.html#csq) with appropriate degrees of freedom (in this case, 1).

Part 2: Code for Model 1

**proc** **mixed** data=HowTO.DataS1 method=ML covtest;

class INDIV;

model Y = x\_avg\_cen x\_dev / htype=**3** ddfm=kr solution cl alpha=**.05**;

**run**;

Part 3: Output

We present just the table of fit statistics produced for the above model:

Fit Statistics

-2 Log Likelihood 957.2

AIC (smaller is better) 965.2

AICC (smaller is better) 965.4

BIC (smaller is better) 979.3

Part 4: Interpretation

The above log-likelihood is compared to that from the model presented in Part A above (model 2). LRT: Model 1-Model 2 = 957.2 - 910.4 = 46.8, df = 1, P < 0.0001, indicating significance between-individual variance in intercept or elevation.

*SAS coding, output, and interpretation (4 parts): REML-method*

Part 1: Background

Extract by hand the -2 res log likelihood from the fit statistics table in the output of two models (the model in Part A is labelled model 2. Model 1 is presented below). Look up difference in -2RLL between the two models in a chi-square table (or web site, http://faculty.vassar.edu/lowry/tabs.html#csq) with appropriate degrees of freedom (in this case, 1).

Part 2: Code for Model 1

**proc** **mixed** data=HowTO.DataS1 method=REML covtest;

class INDIV;

model Y = x\_avg\_cen x\_dev / htype=**3** ddfm=kr solution cl alpha=**.05**;

**run**;

Part 3: Output

We present just the table of fit statistics produced for the above model:

Fit Statistics

-2 Log Likelihood 961.1

AIC (smaller is better) 963.1

AICC (smaller is better) 963.1

BIC (smaller is better) 966.6

Part 4: Interpretation

The above log-likelihood is compared to that from the model presented in Part A above (model 2). LRT: Model 1-Model 2 = 961.1 – 912.9 = 48.2, df = 1, P < 0.0001, indicating significance between-individual variance in intercept or elevation.

EB. LRT-BASED SIGNIFICANCE OF and IN BIVARIATE MMs FOR SCENARIO 5, TABLE 2

*Goal*: LRT-based significance calculation for 1. and 2. in the model described in section “B. BIVARIATE MMs FOR A SCENARIO WHERE TWO PHENOTYPIC ATTRIBUTES WERE BOTH ASSAYED REPEATEDLY AT THE SAME TIME (SCENARIO 3/5; TABLE 2)”. The comparison is between a model where both and are estimated versus one where either 1. or 2. is constrained to zero.

*ASReml coding, and calculation of significance (2 parts)*

Part 1: coding of the constrained .as file (1. = 0)

**ASREML analysis**

**Indiv !A**

**x**

**x\_avg**

**x\_dev**

**y**

**z**

**x\_cen**

**x\_avg\_cen**

**y\_std**

**z\_std**

**C:\Howtodo\DataS1.txt !SKIP 1**

**!MVREMOVE**

**!MAXIT 5000**

**!SLOW 3**

**!EXTRA 5**

**!CONTINUE**

**y z ~ Trait !r Trait.Indiv**

**1 2 1 !STEP 0.001**

**0**

**Trait 0 US !GPUP !S2==1**

**0.7454**

**0.000 0.7764**

**Trait.Indiv 2**

**Trait 0 US !GPZP**

**0.3215**

**0.000 0.2575**

**Indiv**

Part 1: coding of the constrained .as file (2. = 0)

**ASREML analysis**

**Indiv !A**

**x**

**x\_avg**

**x\_dev**

**y**

**z**

**x\_cen**

**x\_avg\_cen**

**y\_std**

**z\_std**

**C:\Howtodo\DataS1.txt !SKIP 1**

**!MVREMOVE**

**!MAXIT 5000**

**!SLOW 3**

**!EXTRA 5**

**!CONTINUE**

**y z ~ Trait !r Trait.Indiv**

**1 2 1 !STEP 0.001**

**0**

**Trait 0 US !GPZP !S2==1**

**0.7454**

**0.000 0.7764**

**Trait.Indiv 2**

**Trait 0 US !GPUP**

**0.3215**

**0.000 0.2575**

**Indiv**

Part 2: Calculation of significance of and

1. **Loglikelihood Unconstrained = -434.832 (asr file of unconstrained model)**
2. **Loglikelihood Constrained (1.**  = 0**) = -441.440 (asr file of constrained model – not printed)**

**2. Loglikelihood Constrained (2.**  = 0**) = -434.835 (asr file of constrained model – not printed)**

1. **LRT: *χ21=13.216, P<0.001; χ21=0.006, P=0.938***

*R coding, output, and interpretation (3 parts)*

Part 1: coding of the constrained model (1. = 0)

yz.bivar1.mcmc<-

MCMCglmm(cbind(y,z)~(trait-1),

random=~idh(trait):indiv,rcov=~us(trait):units,

family=c("gaussian","gaussian"), prior=prior.bivar,

data=set1.data,verbose=FALSE)

#return the DIC value:

summary(yz.bivar1.mcmc)$DIC

Part 2: coding of the constrained model (2. = 0)

yz.bivar2.mcmc<-

MCMCglmm(cbind(y,z)~(trait-1),

random=~us(trait):indiv,rcov=~idh(trait):units,

family=c("gaussian","gaussian"), prior=prior.bivar,

data=set1.data,verbose=FALSE)

#return the DIC value:

summary(yz.bivar2.mcmc)$DIC

Part 3: Determining statistical support for or against the unconstrained model

1. DIC Unconstrained: 1733.672 (from B)
2. DIC Constrained 1: 1740.552 (from Part 1, this section)
3. DIC Constrained 2: 1731.719 (from Part 2, this section)
4. DIC Unconstrained – DIC Constrained 1: -6.88
5. DIC Unconstrained – DIC Constrained 2: 1.953

Based on the differences, the unconstrained model was substantially better supported by the data than a model (1) in which the between-individual covariances were constrained to zero. However, the unconstrained model is less well supported than one in which the within-individual covariances are constrained to zero.

*SAS 9.2 coding, output, and interpretation* *(4 parts): ML-method*

Part 1. Background

Tests for significance of covariance parameters in SAS using the LRT also require running a constrained and unconstrained model. Constraining the covariances involves modifying the G or R matrix to eliminate the covariance of interest. In SAS, this is accomplished by altering the “Type” statement in either the random statement, thus altering the G matrix and which between-individual covariances are included, or the repeated statement, thus altering the R matrix and which within-individual covariances are included.

Part 2: coding of the constrained model (1. = 0)

Here the G matrix is constrained to have 0 covariance by putting “TYPE = VC” in the random statement:

**Proc** **mixed** data=HowTO.DataS1tall METHOD=ML COVTEST;

class INDIV TRAIT OBS;

model RESPONSE=TRAIT /noint solution;

random TRAIT / Type=vc sub=INDIV g;

repeated TRAIT OBS /type=UN@CS sub=INDIV r;

**run**;

Part 3: coding of the constrained model (2. = 0)

Here the R matrix is constrained by removing the repeated measure and using “TYPE=VC” in the repeated statement.

**Proc** **mixed** data=HowTO.DataS1tall METHOD=ML COVTEST;

class INDIV TRAIT OBS;

model RESPONSE=TRAIT /noint solution;

random TRAIT / Type=un sub=INDIV g;

repeated /type=vc grp = TRAIT sub=INDIV r;

**run**;

Part 4: Determining statistical support for or against the unconstrained models

1. -2 Loglikelihood Unconstrained = 1781.0 (From model in Part B above but run with ML instead of REML)
2. -2 Loglikelihood Constrained (1. = 0) = 1794.7 (in table “Fit Statistics”, not shown)

2. -2 Loglikelihood Constrained (2. = 0) = 1781.0 (in table “Fit Statistics”, not shown)

1. LRT: *χ21=13.7, P<0.001; χ21=0.00, P=1.00,* indicating significant between-individual between-trait covariance but no evidence for any within-individual between-trait covariance.

*SAS 9.2 coding, output, and interpretation* *(4 parts): REML-method*

Part 1. Background

Tests for significance of covariance parameters in SAS using the LRT also require running a constrained and unconstrained model. Constraining the covariances involves modifying the G or R matrix to eliminate the covariance of interest. In SAS, this is accomplished by altering the “Type” statement in either the random statement, thus altering the G matrix and which between-individual covariances are included, or the repeated statement, thus altering the R matrix and which within-individual covariances are included.

Part 2: coding of the constrained model (1. = 0)

Here the G matrix is constrained to have 0 covariance by putting “TYPE = VC” in the random statement:

**Proc** **mixed** data=HowTO.DataS1tall METHOD=REML COVTEST;

class INDIV TRAIT OBS;

model RESPONSE=TRAIT /noint solution;

random TRAIT / Type=vc sub=INDIV g;

repeated TRAIT OBS /type=UN@CS sub=INDIV r;

**run**;

Part 3: coding of the constrained model (2. = 0)

Here the R matrix is constrained by removing the repeated measure and using “TYPE=VC” in the repeated statement.

**Proc** **mixed** data=HowTO.DataS1tall METHOD=REML COVTEST;

class INDIV TRAIT OBS;

model RESPONSE=TRAIT /noint solution;

random TRAIT / Type=un sub=INDIV g;

repeated /type=vc grp = TRAIT sub=INDIV r;

**run**;

Part 4: Determining statistical support for or against the unconstrained models

1. -2 Res Loglikelihood Unconstrained = 1784.9 (From model in Part B above)
2. -2 Res Loglikelihood Constrained (1. = 0) = 1798.4 (in table “Fit Statistics”, not shown)
3. -2 Res Loglikelihood Constrained (2. = 0) = 1784.9 (in table “Fit Statistics”, not shown)
4. LRT: *χ21=13.5, P<0.001; χ21=0.00, P=1.00,* indicating significant between-individual between-trait covariance but no evidence for any within-individual between-trait covariance.

EC. LRT-BASED SIGNIFICANCE OF IN BIVARIATE MMs FOR SCENARIO 4, TABLE 2

*Goal*: LRT-based significance calculation for in the model described in section “C. BIVARIATE MMs FOR A SCENARIO WHERE TWO PHENOTYPIC ATTRIBUTES WERE BOTH ASSAYED REPEATEDLY BUT NEVER AT THE SAME TIME (SCENARIO 4; TABLE 2)”. The comparison is between a model where both is estimated versus one where is constrained to zero.

*ASReml coding, and calculation of significance (2 parts)*

Part 1: coding of the constrained .as file

**ASREML analysis**

**Indiv !A**

**y**

**z**

**y\_std**

**z\_std**

**C:\Howtodo\DataS2.txt !SKIP 1**

**!MVREMOVE**

**!MAXIT 5000**

**!SLOW 3**

**!EXTRA 5**

**!CONTINUE**

**y z ~ Trait !r Trait.Indiv**

**1 2 1 !STEP 0.001**

**0**

**Trait 0 US !GPZP !S2==1**

**0.7454**

**0.000 0.7764**

**Trait.Indiv 2**

**Trait 0 US !GPZP**

**0.3215**

**0.000 0.2575**

**Indiv**

Part 2: Calculation of significance of

1. **Loglikelihood Unconstrained = -434.835 (asr file of unconstrained model)**
2. **Loglikelihood Constrained (** = 0**) = -441.499 (asr file of constrained model – not printed)**
3. **LRT: *χ21=13.328, P<0.001***

*R coding, output, and interpretation (2 parts)*

Part 1: coding of the constrained model ( = 0)

yz.bivarDalt.mcmc<-MCMCglmm(cbind(y,z)~(trait-1),

random=~idh(trait):indiv,rcov=~idh(trait):units,

family=c("gaussian","gaussian"), prior=prior.bivar,

nitt=130000,thin=100,burnin=30000, data=set2.data,verbose=FALSE)

#return the DIC value:

summary(yz.bivarDalt.mcmc)$DIC

Part 3: Determining statistical support for or against the unconstrained model

1. DIC Unconstrained: 1731.828 (from C)
2. DIC Constrained 1: 1738.257 (from Part 1, this section)
3. DIC Unconstrained – DIC Constrained 1: -6.88

Based on the difference in DIC values (-6.429), the unconstrained model was substantially more well supported by the data than a model in which the between-individual covariances were constrained to zero.

*SAS 9.2 coding, output, and interpretation* *(2 parts): ML-method*

Part 1. Coding of the constrained model (1. = 0)

**Proc** **mixed** data=HowTO.DataSets2tall METHOD=ML COVTEST;

class INDIV TRAIT;

model RESPONSE=TRAIT /noint solution;

random intercept/ Type=vc grp=Trait sub=INDIV g;

repeated / Type=VC grp=trait sub=INDIV;

**run**;

Part 2: Determining statistical support for or against the unconstrained model

1. -2\*Loglikelihood Unconstrained = 1781.0 (from ML version of model in Part C above)
2. -2\*Loglikelihood Constrained ( = 0) = 1794.6 – from “Fit Statstics”, not printed)
3. LRT: *χ21=13.6, P<0.001,* indicating significant between-individual, between-trait covariance.

*SAS 9.2 coding, output, and interpretation* *(2 parts): REML-method*

Part 1. Coding of the constrained model (1. = 0)

**Proc** **mixed** data=HowTO.DataSets2tall METHOD=REML COVTEST;

class INDIV TRAIT;

model RESPONSE=TRAIT /noint solution;

random intercept/ Type=vc grp=Trait sub=INDIV g;

repeated / Type=VC grp=trait sub=INDIV;

**run**;

Part 2: Determining statistical support for or against the unconstrained model

1. -2\*Res Loglikelihood Unconstrained = 1784.9 (from model in Part C above)
2. -2\*Res Loglikelihood Constrained ( = 0) = 1798.3 – from “Fit Statstics”, not printed)
3. LRT: *χ21=13.4, P<0.001,* indicating significant between-individual, between-trait covariance.

ED. LRT-BASED SIGNIFICANCE OF DIFFERENCES IN REPEATABILITY BETWEEN TWO DATASETS

*Goal*: LRT-based significance calculation for the difference in repeatability between females and males for the model described in section “D. BIVARIATE MMs TO ESTIMATE REPEATABILITY FOR TWO DATASETS SIMULTANEOUSLY”. The comparison is between the unconstrained model where both and are both estimated versus one where they are constrained to be the same value (i.e. = ).

*ASReml coding, and calculation of significance (2 parts)*

Part 1: coding of the constrained .as file

**ASREML analysis**

**Indiv !A**

**y**

**z**

**y\_std**

**z\_std**

**C:\Howtodo\DataS3.txt !SKIP 1**

**!MVREMOVE**

**!MAXIT 5000**

**!SLOW 3**

**!EXTRA 5**

**!CONTINUE**

**y\_std z\_std ~ Trait !r Trait.Indiv**

**1 2 1 !STEP 0.001**

**0**

**Trait 0 US !GPZP !S2==1**

**0.7454**

**0.000 0.7764**

**Trait.Indiv 2**

**Trait 0 US !GPZP !=aba**

**0.3215**

**0.000 0.2575**

**Indiv**

Part 2: Calculation of significance of attribute-specific repeatabilities

1. **Loglikelihood Unconstrained = -205.755 (asr file of unconstrained model)\***
2. **Loglikelihood Constrained (** = **) = -205.755 (asr file of constrained model – not printed)\***
3. **LRT: *χ21=0, P<1.000***

**\*Note that the values of repeatability for *y* and *z* are so similar that ASReml returns the same Loglikelihood for both the constrained and unconstrained model**

*R coding, output and interpretation (3 parts)*

Part 1: Background

Unlike with ASReml we are not aware of methods by which to directly constrain elements of variance-covariance matrices in MCMCglmm. Thus we will instead compare the fit of the model from Part D with a univariate model in which only a single *Vind* and *Ve* (and thus the same repeatability for males and females). Recall that separate repeatabilities were estimated from the following model:

yz.bivarE.mcmc<-MCMCglmm(cbind(y\_std,z\_std)~(trait-1),

random=~idh(trait):indiv,rcov=~idh(trait):units,

family=c("gaussian","gaussian"), prior=prior.bivar,

nitt=130000,thin=100,burnin=30000, data=set3.data,verbose=FALSE)

Part 2: coding of the constrained model

To fit a univariate model we must first resort the data to have a single variable:

set3.new<-as.data.frame(rbind(as.matrix(subset(set3.data, y!="NA")[,c(1,4)]),

as.matrix(subset(set3.data, z!="NA")[,c(1,5)])))

names(set3.new)<-c("indiv","resp")

We then fit a univariate model:

#use an inverse gama as in the case of yz.bivarE.mcmc;

#please refer to earlier caveats about priors

prior.univar<-list(R=list(V=1, nu=0.002),G=list(G1=list(V=1, nu=0.002)))

yz.bivarEalt.mcmc<-MCMCglmm(resp~1,

random=~indiv, family="gaussian", prior=prior.univar,

nitt=130000,thin=100,burnin=30000, data=set3.data,verbose=FALSE)

Part 3: Determining statistical support for or against differing variance components by dataset

1. DIC differing variance components: 1264.73 (from D)
2. DIC same variance components: 633.64 (from Part 2, this section)
3. DIC differing – DIC same: 631.09

Based on the difference in DIC values, there is considerable support that the variance components do not differ between data sets

*SAS coding, output, and interpretation (4 parts): ML-method*

Part 1: Background

Tests for differences in repeatability between two or more groups using the likelihood ratio test can be achieved in SAS using a univariate approach. In the case of measuring two traits in two subclasses of individuals, SAS can treat the analysis as if it were one trait but controlling for trait identity. Using dataset S3 modified to allow bivariate modeling, e.g., by stacking all measurements (“Stdresp”) in one column and coding a second column (“Trait”) to identify trait, we then do two univariate models of the standardized response, one with no distinction between the identity of the individuals, and one in which the random effect of individual is split by trait into two groups. We have included “trait” as well in the fixed effects to illustrate that one might want to account for any mean differences between the traits, but since we are using standardized values, this is not necessary in this particular case. Note that this procedure can be used to test differences in repeatability between groups in the same trait (e.g., differences in repeatability of parental care between the sexes, Westneat et al. 2011).

Part 2: Code

**proc** **mixed** data=HowTO.DataS3tall METHOD=ML CL ALPHA=**.05** COVTEST;

class Indiv Trait;

model Stdresp = trait / HTYPE=**3** DDFM=KR SOLUTION CL ALPHA=**.05**;

random int/subject=Indiv;

**run**;

**proc** **mixed** data=HowTO.DataS3tall METHOD=ML CL ALPHA=**.05** COVTEST;

class Indiv Trait;

model Stdresp = trait / HTYPE=**3** DDFM=KR SOLUTION CL ALPHA=**.05**;

random int/subject=Indiv Group = trait;

**run**;

Part 3: Output

We provide several parts of the output for both models, but the essential bit for LRT is the -2 log likelihood in the table of fit statistics.

Model 1

Covariance Parameter Estimates

Standard Z

Cov Parm Subject Estimate Error Value Pr > Z Alpha Lower Upper

Intercept Indiv 0.3721 0.07082 5.25 <.0001 0.05 0.2646 0.5617

Residual 0.6239 0.04412 14.14 <.0001 0.05 0.5457 0.7203

Fit Statistics

-2 Log Likelihood 1321.2

AIC (smaller is better) 1329.2

AICC (smaller is better) 1329.3

BIC (smaller is better) 1339.7

Solution for Fixed Effects

Standard

Effect Trait Estimate Error DF t Value Pr > |t| Alpha Lower Upper

Intercept -214E-18 0.09969 100 -0.00 1.0000 0.05 -0.1978 0.1978

Trait Y 2.71E-16 0.1410 100 0.00 1.0000 0.05 -0.2797 0.2797

Trait Z 0 . . . . . . .

Model 2

Fit Statistics

-2 Log Likelihood 1321.2

AIC (smaller is better) 1331.2

AICC (smaller is better) 1331.4

BIC (smaller is better) 1344.3

Part 4: Interpretation

The repeatability in model 1 over all individuals and thus over both traits is very similar to the repeatabilities obtained in model two that are split between the two traits, and those repeatabilities are extremely similar to each other. The -2 residual log likelihoods are identical for the two models, so the LRT is: difference in -2RLL = chi-square = 0, df = 1, P = 1.0.

*SAS coding, output, and interpretation (4 parts): REML-method*

Part 1: Background

Tests for differences in repeatability between two or more groups using the likelihood ratio test can be achieved in SAS using a univariate approach. In the case of measuring two traits in two subclasses of individuals, SAS can treat the analysis as if it were one trait but controlling for trait identity. Using dataset S3 modified to allow bivariate modeling, e.g., by stacking all measurements (“Stdresp”) in one column and coding a second column (“Trait”) to identify trait, we then do two univariate models of the standardized response, one with no distinction between the identity of the individuals, and one in which the random effect of individual is split by trait into two groups. We have included “trait” as well in the fixed effects to illustrate that one might want to account for any mean differences between the traits, but since we are using standardized values, this is not necessary in this particular case. Note that this procedure can be used to test differences in repeatability between groups in the same trait (e.g., differences in repeatability of parental care between the sexes, Westneat et al. 2011).

Part 2: Code

**proc** **mixed** data=HowTO.DataS3tall METHOD=REML CL ALPHA=**.05** COVTEST;

class Indiv Trait;

model Stdresp = trait / HTYPE=**3** DDFM=KR SOLUTION CL ALPHA=**.05**;

random int/subject=Indiv;

**run**;

**proc** **mixed** data=HowTO.DataS3tall METHOD=REML CL ALPHA=**.05** COVTEST;

class Indiv Trait;

model Stdresp = trait / HTYPE=**3** DDFM=KR SOLUTION CL ALPHA=**.05**;

random int/subject=Indiv Group = trait;

**run**;

Part 3: Output

We provide several parts of the output for both models, but the essential bit for LRT is the -2 res log likelihood in the table of fit statistics.

Model 1

Covariance Parameter Estimates

Standard Z

Cov Parm Subject Estimate Error Value Pr > Z Alpha Lower Upper

Intercept Indiv 0.3822 0.07296 5.24 <.0001 0.05 0.2716 0.5778

Residual 0.6239 0.04412 14.14 <.0001 0.05 0.5457 0.7203

Fit Statistics

-2 Res Log Likelihood 1326.8

AIC (smaller is better) 1330.8

AICC (smaller is better) 1330.8

BIC (smaller is better) 1336.0

Solution for Fixed Effects

Standard

Effect Trait Estimate Error DF t Value Pr > |t| Alpha Lower Upper

Intercept -213E-18 0.1007 98 -0.00 1.0000 0.05 -0.1998 0.1998

Trait Y 2.89E-16 0.1424 98 0.00 1.0000 0.05 -0.2826 0.2826

Trait Z 0 . . . . . . .

Model 2

Fit Statistics

-2 Res Log Likelihood 1326.8

AIC (smaller is better) 1332.8

AICC (smaller is better) 1332.8

BIC (smaller is better) 1340.6

Part 4: Interpretation

The repeatability in model 1 over all individuals and thus over both traits is very similar to the repeatabilities obtained in model two that are split between the two traits, and those repeatabilities are extremely similar to each other. The -2 residual log likelihoods are identical for the two models, so the LRT is: difference in -2RLL = chi-square = 0, df = 1, P = 1.0.

**Supplementary Table S1.** Eight useful types of univariate GLMMs with random intercepts (and slopes) for the factor “Individual”. For each linear equation we give a verbal description of the model and the context in which it may be applied.

|  |  |  |
| --- | --- | --- |
| Model number and equation | Model description | Application |
| S1. | Univariate LMM with random intercepts for individual () | Calculation of repeatability (*r*)1 of the raw data: individual () and residual () variances (hence *r*) are estimated without controlling for any fixed effects |
| S2. | Univariate LMM with random intercepts for individual () with a between-individual fixed effect () | Calculation of the average within-class repeatability for a set of classes of individual1,2: represents the between-individual variance not accounted for by the between-individual fixed effect |
| S3. | Univariate LMM with random intercepts for individual () with each individual’s average value of a covariate fitted as a between-individual fixed effect () | Calculation of repeatability1 of the raw data while avoiding ‘pseudo-repeatability’3: represents the between-individual variance not accounted for by a between-individual fixed effect |
| S4. | Univariate LMM with random intercepts for individual () with a within-individual fixed effect () | Calculation of repeatability1,2 while controlling for variation in environmental conditions () across repeated trials, applicable only to cases where all individuals experience the same set of conditions4. represents the within-individual variance not accounted for by the within-individual fixed effect |
| S5. | Univariate LMM with random intercepts for individual () with the dependence of on fixed effect estimated within () and between individuals () using ‘within-subject centering’4 | Calculation of repeatability1,2 while controlling for variation in conditions across repeated trials and avoiding ‘pseudo-repeatability’3:  represents the between-individual variance not accounted for by the between-individual component of the fixed effect (), represents the within-individual variance not accounted for by the within-individual component of the fixed effect ()4. |
| S6. | Alternative notation of Model S5, where the dependence of on fixed effect is split between the within-individual effect () and the difference between the between- and within-individual effect | Assessment of differential dependence of on within versus between individuals is estimated by a single parameter. Whenever , individuals are distributed non-randomly over the covariate with respect to their intercept5. |
| S7. | Univariate LMM with random intercepts for individual () with a within-individual fixed effect (), where random slopes () are fitted around the population-average slope () of the dependence of on | Modeling the between-individual variance (hence repeatability) as a function of a covariate (); estimates individual variance in plasticity () and its covariance () with 6. Applicable only to cases where all individuals experience the same set of conditions4. |
| S8. | Univariate LMM with random intercepts for individual () with the dependence of on fixed effect estimated within () and between individuals () using ‘within-subject centering’4, where random slopes () are fitted around the population-average slope () of the dependence of on () | As Model S7 but applicable to situations where individuals differ in average value of the covariate ()4,6,7. |
| S9. | As model S8, with an interactive effect of dependence of on fixed effect within × between individuals () | As model S8; represents a situation where individuals are distributed non-randomly over the covariate with respect to their level of plasticity8. |

**Supplementary Table S1, footnotes**

1Repeatability (*r*) is calculated as: .

2The model assumes that neither nor vary as a function of the fixed effect; for categorical between-individual fixed effects (e.g. sex), estimated values of repeatability1 represent the average within-class repeatability in cases where sample sizes are equal for all classes.

3Pseudo-repeatability occurs when between-individual variation in a fixed effect influencing the phenotype *within* individuals is not appropriately controlled for (Westneat et al. In Press). For example, parental birds increase their provisioning rate increases with nestling age. Values of (hence repeatability1) would be over-estimated if the experimental design allowed the average nestling age () over all observations of an individual (*j*) to vary between individual parents. Inclusion of into the model would prevent inflated values of repeatability.

4Fixed effects may vary both within and between individuals, and the dependence () of response variables on fixed effects may also vary within () versus between () individuals. Between- versus within-individual dependence of the response variable on a focal fixed effect therefore needs to be teased apart (van de Pol & Wright 2009). This is done by calculating the average covariate value for the individual (), the deviation from each value for each measurement (), and then fitting these two predictor variables (instead of ) into the model. Only in special cases where does not vary between individuals (e.g. all individuals were assayed at the same set of ages, or times of day, as part of the experimental design), varies solely within individuals and the dependence represents an unbiased estimate of .

5For illustration see Fig.1b in van de Pol & Wright (2009). In the context of behavioural traits measured in natural situations as part of observational studies, and where the covariate represents an environmental condition (e.g. density during the measurement of the phenotype), whenever there is ‘personality-related sampling bias’, because certain types of individuals are more likely to occur in certain types of environment, where personality is defined as the intercept of an individual’s behavioural reaction norm (Dingemanse *et al.* 2010c).

6This so-called ‘random regression’ models the between-individual variance as a function of a covariate, while simultaneously modeling either homogeneous (; Models S7-S9) or heterogeneous residual errors for different values of the covariate (not shown), for further details see Schaeffer (2004). Repeatability explicitly varies as function of the covariate, where the classic equation1 represents the repeatability for the specific condition where (Model S7) or (Models S8-S9).

7Note that uncounted non-linearity of the dependence of on fixed effect may result in as a statistical artifact

8Expanding upon the example presented in footnote 5, represents a situation where individuals are non-randomly distributed over the environment with respect to their level of plasticity. Note that uncounted non-linearity of the dependence of on fixed effect may result in .

**Supplementary Table S2.** A variety of multivariate methods are available to elucidate covariance patterns both within and among populations. The listed methods are some which have been used within the quantitative genetics literature but less broadly elsewhere in evolutionary ecology. While often formulated for genetic covariance matrices the approaches listed can be used with covariances matrices derived from MMs with strictly phenotypic data, although certain assumptions may have to be made in such cases.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Method | Question(s) that can be asked | Within groups | Between groups | Software packages | Refs. | |
| Mantel Test | Are the elements of two covariance/correlation matrices similar | No | Yes | Most software packages | Mantel (1967) | |
| Common Principal Components | Are the eigenvalues and eigenvectors of two matrices similar | No | Yes | Purpose built software | Phillips & Arnold (1989) | |
| Random Skewers Analysis | Do the covariances for two matrices result in similar evolutionary trajectories? | No | Yes | - | Cheverud & Marroig (2007) | |
| Element-by-Element Comparison | Do specific elements differ between matrices? | No | Yes | Most software packages | Dingemanse *et al.* (2009) | |
| Conditional Evolvability Analysis | Do the variances and covariances of a matrix constrain the ability of a population to respond to evolutionary pressures? Do two matrices differ in this regard? | Yes | Yes | - | Hansen & Houle (2008) |
| Confirmatory Factor Analysis | Are specific hypotheses regarding (co)variances supported by the data? | Yes | Yes | Most software packages | Dochtermann & Jenkins (2007); Dingemanse et al. (2010a) |
| Multivariate Analysis of Variance | Do the phenotypic centroids differ based on *a priori* considered causal factors? | Yes | Yes | Most software packages†† | Roff (2002) |

† Although many statistical packages allow multivariate analyses of variances, fewer do so using a mixed model approach.

**Figure S1.** We illustrate here graphically how between- and within-individual variance components are separated by plotting seven measurements of aggressiveness (y-axis) for five individuals (numbered) whose behaviour was assayed over a range of conspecific densities (x-axis). Grey lines represent the average phenotypic value of each individual; the variance among lines represents the between-individual variance (). The variance in within-individual deviation from individual means represents the within-individual variance (). Each individual is sampled only in a restricted range of environments. The relationship between phenotype and environment differs between (positive relationship; ) versus within individuals (negative relationship; ), as demonstrated by the difference between the slopes for individual responses (grey lines) and the population trend (black line). The values in bold face represent each individual’s average value for *x* and *y*.



**Figure S2.** Three models of syndrome structure developed based on a priori hypotheses of behavioural syndrome structure in Merriam’s kangaroo rats proposed by Dochtermann & Jenkins (2007), and detailed in the main text. The measured variables are representedin rectangular boxes. Underlying causal connections resulting in syndrome structure are presented in ovals. These connections can be modelled with latent variables (Dingemanse *et al.* 2010a) or by placing specific constraints on the multivariate MM’s covariance matrix (Eqns. S9b-d).

**C:\Users\dochtermann\Dropbox\Working\HowToPaper\FigS1 - Copy.tif**

Supplementary References

Biro, P. A., Beckmann, C. & Stamps, J. A. (2010) Small within-day increases in temperature affects boldness and alters personality in coral reef fish. *Proceedings of the Royal Society of London Series B*, **277,** 71-77.

Biro, P. A. & Dingemanse, N. J. (2009) Sampling bias resulting from animal personality. *Trends in Ecology and Evolution*, **24,** 66-67.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H. & White, J. S. S. (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, **24,** 127-135.

Brommer, J. E., Rattiste, K. & Wilson, A. J. (2008) Exploring plasticity in the wild: laying date-temperature reaction norms in the common gull *Larus canus*. *Proceedings of the Royal Society of London Series B*, **275,** 687-693.

Browne, W. J., McCleery, R. H., Sheldon, B. C. & Pettifor, R. A. (2007) Using cross-classified multivariate mixed response models with application to life history traits in great tits (*Parus major*). *Statistical Modelling*, **7,** 217-238.

Cheverud, J. M. & Marroig, G. (2007) Comparing covariance matrices: Random skewers method compared to the common principal components model. *Genetics and Molecular Biology*, **30,** 461-469.

Cleasby, I. R. & Nakagawa, S. (2011) Neglected biological patterns in the residuals. *Behavioural Ecology and Sociobiology*, **65,** 2361-2371.

Conover, D. O. & Schultz, E. T. (1995) Phenotypic similarity and the evolutionary significance of countergradient variation. *Trends in Ecology and Evolution*, **10,** 248-252.

Davis, J. A., Spaeth, J. L. & Huson, C. (1961) A technique for analyzing the effects of group composition. *American Sociological Review*, **26,** 215-225.

Dingemanse, N. J., Barber, I., Wright, J. & Brommer, J. E. (2012a) Quantitative genetics of behavioural reaction norms: genetic correlations between personality and behavioural plasticity vary across stickleback populations. *Journal of Evolutionary Biology*, **25,** 485-496.

Dingemanse, N. J., Bouwman, K. M., van de Pol, M., van Overveld, T., Patrick, S. C., Matthysen, E. & Quinn, J. L. (2012b) Variation in personality and behavioural plasticity across four populations of the great tit *Parus major*. *Journal of Animal Ecology*, **81,** 116-126.

Dingemanse, N. J., Dochtermann, N. A. & Wright, J. (2010a) A method for exploring the structure of behavioural syndromes to allow formal comparison within and between datasets. *Animal Behaviour*, **79,** 439-450.

Dingemanse, N. J., Edelaar, P. & Kempenaers, B. (2010b) Why is there variation in glucocorticoids? *Trends in Ecology and Evolution*, **25,** 261-262.

Dingemanse, N. J., Kazem, A. J. N., Réale, D. & Wright, J. (2010c) Behavioural reaction norms: where animal personality meets individual plasticity. *Trends in Ecology and Evolution*, **25,** 81-89.

Dingemanse, N. J., van der Plas, F., Wright, J., Réale, D., Schrama, M., Roff, D. A., van der Zee, E. & Barber, I. (2009) Individual experience and evolutionary history of predation affect expression of heritable variation in fish personality and morphology. *Proceedings of the Royal Society of London Series B*, **276,** 1285-1293.

Dochtermann, N. A. (2010) Behavioral syndromes: carry-over effects, false discovery rates and *a priori* hypotheses. *Behavioral Ecology*, **21,** 437-439.

Dochtermann, N. A. (2011) Testing Cheverud's conjecture for behavioral correlations and behavioral syndromes. *Evolution*, **65,** 1814-1820.

Dochtermann, N. A. & Jenkins, S. H. (2007) Behavioural syndromes in Merriam's kangaroo rats (*Dipodomys merriami*): a test of competing hypotheses. *Proceedings of the Royal Society of London Series B*, **274,** 2343-2349.

Dochtermann, N. A. & Roff, D. A. (2010) Applying a quantitative genetics framework to behavioural syndrome research. *Philosophical Transactions of the Royal Society of London Series B*, **365,** 4013-4020.

Enders, C. K. & Tofighi, D. (2007) Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, **12,** 121-138.

Garamszegi, L. Z., Calhim, S., Dochtermann, N., Hegyi, G., Hurd, P. L., Jorgensen, C., Kutsukake, N., Lajeunesse, M. J., Pollard, K. A., Schielzeth, H., Symonds, M. R. E. & Nakagawa, S. (2009) Changing philosophies and tools for statistical inferences in behavioral ecology. *Behavioral Ecology*, **20,** 1363-1375.

Gilchrist, G. W. (1996) A quantitative genetic analysis of thermal sensitivity in the locomotor performance curve of *Aphidius ervi*. *Evolution*, **50,** 1560-1572.

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J. & Thompson, R. (2006) *ASreml user guide. Release 1.0*. VSN International, Hemel Hempstead, U.K.

Hadfield, J. D. (2010) MCMC methods for multi-response generalized linear mixed models: The MCMCglmm *R* package. *Journal of Statistical Software*, **33,** 1-22.

Hansen, T. F., Armbruster, W. S., Carlson, M. L. & Pelabon, C. (2003) Evolvability and genetic constraint in *Dalechampia* blossoms: Genetic correlations and conditional evolvability. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution*, **296B,** 23-39.

Hansen, T. F. & Houle, D. (2008) Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology*, **21,** 1201-1219.

Koolhaas, J. M., De Boer, S. F., Coppens, C. M. & Buwalda, B. (2010) Neuroendocrinology of coping styles: Towards understanding the biology of individual variation. *Frontiers in Neuroendocrinology*, **31,** 307-321.

Kreft, I. G. G., Deleeuw, J. & Aiken, L. S. (1995) The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, **30,** 1-21.

Laland, K. N., Odling-Smee, F. J. & Feldman, M. W. (1999) Evolutionary consequences of niche construction and their implications for ecology. *Proceedings of the National Academy of Sciences of the United States of America*, **96,** 10242-10247.

Lande, R. (1979) Quantitative genetics analysis of multivariate evolution, applied to brain:body size allometry. *Evolution*, **33,** 402-416.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D. & Schabenberger, O. (2006) *SAS for mixed models, 2nd edition*. SAS Institute Inc., Carry, NC.

Longford, N. T. (1989) Contextual effects and group means. *Multilevel Modelling Newletter*, **1,** 5-11.

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

Mantel, N. (1967) Detection of disease clustering and a generalized regression approach. *Cancer Research*, **27,** 209-&.

Nakagawa, S. & Schielzeth, H. (2010) Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol.Rev.*, **85,** 935-956.

Paccagnella, O. (2006) Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. *Evaluation Review*, **30,** 66-85.

Phillimore, A. B., Hadfield, J. D., Jones, O. R. & Smithers, R. J. (2010) Differences in spawning date between populations of common frog reveal local adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, **107,** 8292-8297.

Phillips, P. C. & Arnold, S. J. (1989) Visualizing multivariate selection. *Evolution*, **43,** 1209-1222.

Pinheiro, J. C. & Bates, D. M. (2000) *Mixed effect models in S and S-PLUS*. Springer, New York.

Plewis, I. (1989) Comment on "centering" predictors in multilevel analysis. *Multilevel Modelling Newletter*, **1,** 6-11.

Rasbash, J., Steele, F., Browne, W. & Prosser, B. (2005) *A User's Guide to MLwiN - Version 2.0*. Centre for Multilevel Modelling, University of Bristol, London.

Raudenbush, S. W. (1989a) A response to Longford and Plewis. *Multilevel Modelling Newletter*, **1,** 8-11.

Raudenbush, S. W. (1989b) Centering predictors in multilevel analysis: choices and consequences. *Multilevel Modelling Newletter*, **1,** 10-12.

Reznick, D., Nunney, L. & Tessier, A. (2000) Big houses, big cars, superfleas and the costs of reproduction. *Trends in Ecology & Evolution*, **15,** 421-425.

Roff, D. A. (2002) Comparing G matrices: A MANOVA approach. *Evolution*, **56,** 1286-1291.

Rosa, G. J. M., Valente, B. D., de los Campos, G., Wu, X. L., Gianola, D. & Silva, M. A. (2011) Inferring causal phenotypic networks using structural equation modelling. *Genetics Selection Evolution*, **43,** 6.

Scarr, S. & Mccartney, K. (1983) How people make their own environments - a theory of genotype-environment effects. *Child Development*, **54,** 424-435.

Schaeffer, L. R. (2004) Application of random regression models in animal breeding. *Livestock Production Science*, **86,** 35-45.

Scheipl, F., Greven, S. & Kuchenhoff, H. (2008) Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational Statistics & Data Analysis*, **52,** 3283-3299.

Sih, A. & Bell, A. M. (2008) Insights for behavioral ecology from behavioral syndromes. *Advances in the Study of Behavior*, **38,** 227-281.

Snijders, T. A. B. & Bosker, R. J. (1999) *Multilevel analysis - an introduction to basic and advanced multilevel modelling*. Sage, London.

Stamps, J. & Groothuis, T. G. G. (2010) The development of animal personality: relevance, concepts and perspectives. *Biol.Rev.Camb.Philos.Soc2.*, **85,** 301-325.

Stearns, S. C. (1992) *The evolution of life histories*. Oxford University Press, New York.

van de Crommenacker, J., Komdeur, J., Burke, T. & Richardson, D. S. (2011) Spatio-temporal variation in territory quality and oxidative status: a natural experiment in the Seychelles warbler (*Acrocephalus sechellensis*). *Journal of Animal Ecology*, **80,** 668-680.

van de Pol, M. (2012) Quantifying individual variation in reaction norms: how study design affects the accuracy, precision and power of random regression models. *Methods in Ecology and Evolution*, **3,** 268-280.

van de Pol, M. & Verhulst, S. (2006) Age-dependent traits: A new statistical model to separate within- and between-individual effects. *American Naturalist*, **167,** 766-773.

van de Pol, M. & Wright, J. (2009) A simple method for distinguishing within- versus between-subject effects using mixed models. *Animal Behaviour*, **77,** 753-758.

van Noordwijk, A. J. & de Jong, G. (1986) Acquisition and allocation of resources - Their influence on variation in life-history tactics. *American Naturalist*, **128,** 137-142.

Vandermeer, J., Odling-Smee, F. J., Laland, K. N. & Feldman, M. W. (2004) Niche construction - The neglected process in evolution. *Science*, **303,** 472-474.

Visscher, P. M. (2006) A note on the asymptotic distribution of likelihood ratio tests to test variance components. *Twin Research and Human Genetics*, **9,** 490-495.

Westneat, D. F. & Fox, C. W. (2010) *Evolutionary Behavioural Ecology*. Oxford University Press, New York.

Westneat, D. F., Hatch, M. I., Wetzel, D. P. & Ensminger, A. L. (2011) Individual variation in parental care reaction norms: integration of personality and plasticity. *American Naturalist*, **178,** 652-667.

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. *Journal of Animal Ecology*, **79,** 13-26.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A. & Smith, G. M. (2009) *Mixed effect models and extentions in ecology with R*. Springer, New York.