JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.13048

Genetic basis of between-individual and within-individual variance of docility

J. G. A. MARTIN* , E. PIROTTA†, M. B. PETELLE‡ & D. T. BLUMSTEIN§¶

*Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen, UK
†School of Mathematics, Washington State University, Vancouver, WA, USA
‡Department of Zoology and Entomology, University of the Free State Qwaqwa, Phuthaditjhaba, South Africa
§Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA
¶The Rocky Mountain Biological Laboratory, Crested Butte, CO, USA

Keywords:

docility;
double hierarchical generalized linear model;
heritability;
mammals;
Marmota flaviventris;
personality;
predictability;
quantitative genetic;
within-individual variance.

Abstract

Between-individual variation in phenotypes within a population is the basis of evolution. However, evolutionary and behavioural ecologists have mainly focused on estimating between-individual variance in mean trait and neglected variation in within-individual variance, or predictability of a trait. In fact, an important assumption of mixed-effects models used to estimate between-individual variance in mean traits is that within-individual residual variance (predictability) is identical across individuals. Individual heterogeneity in the predictability of behaviours is a potentially important effect but rarely estimated and accounted for. We used 11 389 measures of docility behaviour from 1576 yellow-bellied marmots (*Marmota flaviventris*) to estimate between-individual variation in both mean docility and its predictability. We then implemented a double hierarchical animal model to decompose the variances of both mean trait and predictability into their environmental and genetic components. We found that individuals differed both in their docility and in their predictability of docility with a negative phenotypic covariance. We also found significant genetic variance for both mean docility and its predictability but no genetic covariance between the two. This analysis is one of the first to estimate the genetic basis of both mean trait and within-individual variance in a wild population. Our results indicate that equal within-individual variance should not be assumed. We demonstrate the evolutionary importance of the variation in the predictability of docility and illustrate potential bias in models ignoring variation in predictability. We conclude that the variability in the predictability of a trait should not be ignored, and present a coherent approach for its quantification.

Introduction

Phenotypic variance is a central concept in ecology and evolution as it is the material on which selection can act (Roff, 2002). Phenotypic traits vary among species, among populations and among individuals within populations (Roff, 2002). In addition, for traits expressed multiple times, the phenotype could vary even within an individual (Roff, 2002). Until recently, within-individual variance has been mainly attributed to either environmental plasticity and explained by differences in the environment (Pigliucci, 2005; Nussey *et al.*, 2007), or measurement error and white noise (Pigliucci, 2005; Westneat *et al.*, 2014). Within-individual variance not explained by the general environment is often assumed to be homogeneous across individuals. However, between-individual variation in within-individual variance could exist and be under selection (Hill & Mulder, 2010; Westneat *et al.*, 2014) and thus should not be ignored.

Correspondence: Julien G.A. Martin, Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen, AB24 2TZ, UK.
Tel.: +44 1224 272399; fax: +44 1224 272396;
e-mail: julienmartin@abdn.ac.uk

Existing studies of between-individual variation focus on decomposing the variance in the mean of a trait in its between-individual and within-individual (i.e. residual) components (Kruuk, 2004; Dingemanse & Dochtermann, 2013). In some cases, between-individual variation in environmental plasticity, or individual-by-environment interactions noted IxE, is also estimated (Fig. 1a; e.g. Dingemanse & Dochtermann, 2013; Nussey *et al.*, 2007). However, the residual within-individual variance is always assumed to be the same across individuals (e.g. Wilson *et al.*, 2009; Dingemanse & Dochtermann, 2013). Indeed, the main assumption of mixed-effects models, a statistical approach widely used to estimate between-individual variance both at the phenotypic and genetic level, is that the residual variance is identical across individuals (Pinheiro & Bates, 2000; Wilson *et al.*, 2009; Dingemanse & Dochtermann, 2013). The residual within-individual variance could however differ among individuals at both genetic and phenotypic level. As an example, Fig. 1a illustrates between-individual variation in phenotypic plasticity, whereby the phenotype of three individuals has a different relationship with environmental conditions, that is IxE. Figure 1b shows the residuals of the regression for each individual in Fig. 1a, that is the residual within-individual variance. It is clear from Fig. 1b that the individuals also differ in the amount of variation in their residuals, for example circles are more spread apart than crosses, illustrating between-individual variation in residual within-individual variance. This variance component could have both an environmental and a genetic basis.

The importance of between-individual variation in within-individual variance has been recognized in evolutionary ecology (Mulder *et al.*, 2007) and behavioural ecology (Réale & Dingemanse, 2009; Stamps *et al.*, 2012; Westneat *et al.*, 2013). In agriculture production, the level of variation in yield is of prime importance for growing, processing and consumption of foods, thus leading researchers to investigate the within-genotype variation in yield traits within standardized environments (Hill & Mulder, 2010). In behavioural ecology, the within-individual variance of a trait is often the biggest component of variance (Westneat *et al.*, 2014) suggesting that further investigation is required. Variation in within-individual variance of behaviours has been identified for song repertoire size in birds (Byers & Kroodsma, 2009), antipredatory behaviours (Stamps *et al.*, 2012) and parental care (Westneat *et al.*, 2013). Despite a rising awareness of the importance of variation in within-individual variance in both evolutionary and behavioural ecology (Réale & Dingemanse, 2009; Hill & Mulder, 2010; Stamps *et al.*, 2012; Westneat *et al.*, 2014), empirical studies remain limited.

One of the problems associated with the study of variation in within-individual variance is lexical. For traits measured repeatedly under the same environmental conditions, multiple terms (such as intra-individual variability, individual stability, relative specialization, consistency, predictability or uniformity) have been used as synonyms to refer to within-individual or within-genotype variance (Réale & Dingemanse, 2009; Stamps *et al.*, 2012; Cleasby *et al.*, 2015; Sae-Lim *et al.*, 2015). Following, Cleasby *et al.* (2015), we use predictability to refer to within-individual variance in a trait measured repeatedly in the same environment.

An additional problem for the study of the variation in predictability of behaviours is that the few existing estimates used different, noncomparable approaches (Réale & Dingemanse, 2009; Stamps *et al.*, 2012; Westneat *et al.*, 2013; Cleasby *et al.*, 2015). Some authors

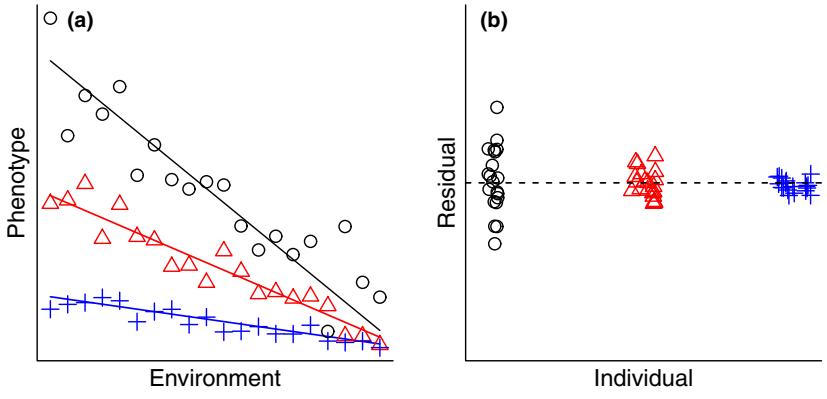

Fig. 1 Example of plasticity (a) and residual within-individual variation (b) for three individuals. Each individual is represented by a different colour and point type. (a) Illustrates between-individual variation in plasticity, with each individual responding differently to environmental changes. The lines represent the average individual response, whereas the points are the observations. (b) Represents the residuals around the mean environmental response for each individual in panel (a). The dotted line is 0. Thus, (b) Illustrates the between-individual variation in residual within-individual variance.

have advocated the use of statistical models to predict an individual's expected behaviour for multiple observations. The standard deviation of the differences between the predictions and the observed behaviours (i.e. model residuals) for each individual then provides an estimate of the variation within an individual (Stamps *et al.*, 2012). This method however suffers from different statistical limitations including potential anticonservatism due to the large amount of uncertainty in the estimation of the residuals not accounted for in subsequent analyses of within-individual variation. Others suggested to use diversity indexes, such as richness, evenness or Shannon Wiener index, to estimate variation in behaviours for each individual (Ram & Gerstorf, 2009). However, the method is limited to discrete behaviours and highly dependent on which diversity index is used. Cleasby *et al.* (2015) presented a method to adequately estimate variation in predictability of a trait using double hierarchical generalized linear models, DHGLM (Lee & Nelder, 2006). DHGLM allows the concurrent estimation of between-individual differences in a trait, or variation in the mean, and in its predictability, while correcting for environmental effects (Cleasby *et al.*, 2015). This type of model is an extension of a mixed-effects model including fixed and random effects on both the mean and the within-individual, that is residual, variance of a trait (Lee & Nelder, 2006; Cleasby *et al.*, 2015). In addition, this approach permits the estimation of comparable parameters of predictability among traits, environments and species (Hill & Mulder, 2010; Cleasby *et al.*, 2015).

Double hierarchical generalized linear models can not only be used to estimate between-individual variance in predictability, but, paired with a quantitative genetic approach, the variance in predictability can also be decomposed into its additive genetic and environmental components (Rönnegård *et al.*, 2010; Sae-Lim *et al.*, 2015). Studies in animal breeding have estimated the additive genetic basis of within-genotype variance for multiple productivity-related traits such as litter size (Hill & Mulder, 2010) or body weight (Sae-Lim *et al.*, 2015). However, the degree to which genetic variation for the predictability of a trait might respond to selection and influence evolution of other traits remains unclear (but see Mulder *et al.*, 2015). In a bivariate scenario, Mulder *et al.* (2015) showed that additive genetic variance in predictability of a linearly selected trait could lead to nonlinear responses for correlated traits. The recent development of a theoretical framework (Hill & Mulder, 2010; Westneat *et al.*, 2014) and of suitable statistical methods (Lee & Nelder, 2006; Cleasby *et al.*, 2015; Mulder *et al.*, 2015) offer a way to advance our understanding of these issues.

To date, the few published estimates of additive genetic variance of predictability are restricted to a captive animal breeding environment (Hill & Mulder, 2010; Mulder *et al.*, 2015; Sae-Lim *et al.*, 2015), except for one study (Mulder *et al.*, 2016). Mulder *et al.* (2016) found significant genetic variation in the variability of fledging weights in wild great tits (*Parus major*). They also showed that the variability in fledging weight was maintained in the population via stabilizing selection. Despite its potential importance, the existence of a genetic basis of predictability in any traits in the wild is poorly understood. Repeated sampling of a large number of pedigreed individuals is needed to apply these models to a wild population. In addition, environmental sampling conditions should be measured carefully to ensure that the variation in predictability is not only due to confounding variation in environmental variation. However, it should be noted that unknown and unmeasured individual- (IxE) or genotype- (GxE) by-environment interactions will also generate between-individual variation in within-individual variance. Estimating individual variance in predictability could thus be a more general approach to investigate IxE as it does not require an environmental covariate and the absence of variability in predictability would indicate an absence of IxE with any environment.

Animal behaviours are easy traits to measure repeatedly in standardized conditions in the wild. Docility, the reaction to being trapped and handled, is part of the shyness-boldness category of temperament traits and reflects an individual's reaction to a risky situation (Réale *et al.*, 2007). Between-individual variation in the mean behavioural response to risky situation has been thoroughly investigated. For example, docility has been found to be repeatable, heritable, and to influence reproduction in several species in the wild (Réale *et al.*, 2007; Petelle *et al.*, 2015). Between-individual variation in predictability has been neglected. However, being predictable or not might be as important as the mean behavioural response when dealing with a risky situation (Stamps *et al.*, 2012).

Using long-term trapping data on yellow-bellied marmots, *Marmota flaviventris*, we estimated between-individual variation in both mean docility and predictability of docility. By means of a quantitative genetic approach, we then implemented an analytical framework that we define as a double hierarchical animal model (DHAM) to decompose the variance of docility and its predictability into their environmental and genetic components.

Materials and methods

Study system

We used behavioural data collected as part of a long-term demographic study on yellow-bellied marmots at the Rocky Mountain Biological Laboratory (RMBL) in Gothic, Colorado, USA (38°77′N, 106°59′W). Marmots are large, facultatively social, subalpine rodents that live in colonies consisting of one or more matrilineal groups

(Frase & Hoffmann, 1980; Armitage, 2014). These colonies usually consist of one adult male, multiple adult females and their offspring. We regularly trap individuals using Tomahawk live traps set at burrow entrances. Once trapped, individuals were weighed, sexed, their ano-genital distance and left hind foot measured, ear tagged, and given an unique dye mark to facilitate identification from afar (Blumstein *et al.*, 2009).

Docility was assessed as an individual's reaction to being trapped and handled (Réale *et al.*, 2007; Petelle *et al.*, 2013). Upon arriving at a trap, individuals were placed inside a cloth-handling bag, and we dichotomously scored (0, 1) whether individuals bit the trap, emitted an alarm call, struggled in the trap or bag, tooth chattered, and whether they hesitated moving from the trap into the handling bag. We summed these scores and subtracted them from a maximum of six to obtain a value between 1 and 6. Thus, a docile individual would have a score of six whereas an individual with a score of one would be nondocile or pugnacious (Réale *et al.*, 2007). We quantified docility over 11 389 trapping events for 1576 individuals of known age and sex from 2002–2014. The mean and range of docility measurements varied widely across individuals (Fig. S1).

Parentage was assigned from DNA samples collected from individuals during their first trapping event. We extracted DNA using a QiaGen QIAamp DNA minikit and genotyped individuals using 12 microsatellites. Alleles were visualized and assigned using GeneMapper 4.1 software (Applied Biosystems). CERVUS (Kalinowski *et al.*, 2007) was then used to assign maternity and paternity using a maximum likelihood method at 95% confidence for the trio. All adult males and females in a colony were used as potential parents. Genetic assignments of maternity confirmed behavioural observations based on juveniles emergence from maternal burrow. For full details on the pedigree reconstruction method, see Olson *et al.* (2012) and Blumstein *et al.* (2010). The pedigree used in these analyses included 1588 individuals with 90% and 83% of the maternal and paternal links known, respectively (see Table S1 for detailed pedigree structure information).

Statistical analysis

The aim of our analysis was to concurrently estimate the between-individual differences in mean docility and its predictability (or residual within-individual variance). Moreover, we wanted to decompose the variability of the mean trait and of its predictability into environmental and genetic components using a quantitative genetic approach. We developed four models differing in their random and residual structure that progressively built towards this goal, whereas allowing us to monitor any change in the estimates as increasing complexity was introduced. Following previous results from this population (Petelle *et al.*, 2013, 2015), we included the following fixed effects in the mean part of all models: trial number, to account for potential habituation to human handling; day of the year, as a proxy for linear seasonal changes in docility; time of day, coded as 0 for AM sampling and 1 for PM sampling; and age, which was a categorical factor with three levels (juveniles, yearlings and adults).

As docility is a discrete ordinal variable varying between 1 and 6, the data were analysed using threshold models (Foulley & Jaffrézic, 2010). This approach assumes that the docility estimates, $\boldsymbol{y}$, result from grouping an underlying continuous variable with a Gaussian distribution, $\boldsymbol{y}^*$, using a probit link function and five cut-off points, $\theta_1 < \theta_2 < \theta_3 < \theta_4 < \theta_5$, to be estimated.

$$\boldsymbol{y} = \begin{cases} 1, & \text{if} \quad \boldsymbol{y}* \leq \theta_1 \\ k, & \text{if} \quad \theta_{k-1} \leq \boldsymbol{y}* < \theta_k \quad \text{for } k = 2,\ 3,\ 4 \text{ and } 5 \\ 6, & \text{if} \quad \theta_5 \leq \boldsymbol{y}* \end{cases} \tag{1}$$

Model 1 was a traditional mixed-effects model, where docility was modelled as a combination of fixed and random effects and variance was assumed to be homogeneous across individuals.

Model 1 can be therefore written as:

$$\boldsymbol{y}* = \boldsymbol{X}\,\boldsymbol{b} + \boldsymbol{M}\,\boldsymbol{yr} + \boldsymbol{Z}\,\boldsymbol{id} + \boldsymbol{e} \tag{2}$$

where $\boldsymbol{yr} \sim N(0,\ \boldsymbol{I}\ \sigma^2_{yr})$, $\boldsymbol{id} \sim N(0,\ \boldsymbol{I}\ \sigma^2_{id})$ and $\boldsymbol{e} \sim N(0,\ \boldsymbol{I}\ \sigma^2_e)$. $\boldsymbol{b}$, $\boldsymbol{yr}$ and $\boldsymbol{pe}$ are the vectors of fixed, random year and random individual identity effects associated with the corresponding incidence matrix $\boldsymbol{X}$, $\boldsymbol{M}$ and $\boldsymbol{Z}$. σ^2_{yr} and σ^2_{id} are the between-year and between-individual random effect variances, respectively. $\boldsymbol{e}$ is the vector of residuals with variance σ^2_e. $\boldsymbol{I}$ is the identity matrix and the notation $\boldsymbol{I}\sigma^2$ signifies that the random effects and the residuals are independently and identically distributed (iid). Constraining the residuals to be iid assumes that there is no variation among individuals in within-individual deviation from the mean (i.e. no variation in predictability).

Similar to model 1, model 2 only considered changes in mean docility. However, this model also included an additive genetic component, fitted as a random effect on the mean (Waldmann, 2009).

Model 2 is therefore equivalent to a classic animal model (Kruuk, 2004), and can be written as:

$$\boldsymbol{y}* = \boldsymbol{X}\,\boldsymbol{b} + \boldsymbol{M}\,\boldsymbol{yr} + \boldsymbol{Z}\,\boldsymbol{pe} + \boldsymbol{Z}\,\boldsymbol{a} + \boldsymbol{e} \tag{3}$$

where building on eqn 2, $\boldsymbol{a} \sim N(0,\ \boldsymbol{A}\ \sigma^2_a)$ and $\boldsymbol{pe} \sim N(0,\ \boldsymbol{I}\ \sigma^2_{pe})$. $\boldsymbol{a}$ and $\boldsymbol{pe}$ are, respectively, the vectors of additive genetic and permanent environment effects. As each individual has an additive genetic as well as a permanent environment effect, both effects have the same design matrix $\boldsymbol{Z}$. σ^2_a is the additive genetic variance of

the mean trait, and $\boldsymbol{A}$ is the additive genetic relationship matrix.

Model 3 was a double hierarchical generalized linear model (DHGLM) (Lee & Nelder, 2006; Cleasby *et al.*, 2015). This included a model for the mean docility (as in model 1), as well as a dispersion part for the residual variance (i.e. predictability). Following SanCristobal-Gaudy *et al.* (1998), the residual variance was modelled on the log-normal scale as a function of the fixed effect of age, with year and individual identity as random effects. We also estimated the correlation between the individual random effect on the mean and on the predictability, using a multivariate Normal distribution. Model 3 can be written as:

$$\begin{aligned} \boldsymbol{y}* &= \boldsymbol{X_m}\,\boldsymbol{b_m} + \boldsymbol{M}\,\boldsymbol{yr_m} + \boldsymbol{Z}\,\boldsymbol{id_m} + \boldsymbol{e} \\ \log(\boldsymbol{\sigma_e^2}) &= \boldsymbol{X_v}\boldsymbol{b_v} + \boldsymbol{Z}\boldsymbol{id_v} \end{aligned} \quad (4)$$

where $\boldsymbol{e} \sim N(0,\ Diag\ [\boldsymbol{\sigma_e^2}])$, $\begin{bmatrix} \boldsymbol{pe_m} \\ \boldsymbol{pe_v} \end{bmatrix} \sim N\left(\begin{matrix} 0 \\ \\ 0 \end{matrix}, \Sigma_{id} \otimes \boldsymbol{I}\right)$ and $\Sigma_{pe} = \begin{bmatrix} \sigma^2_{id_m} & \sigma_{id_m}\sigma_{id_{v,exp}}\rho_{id} \\ \sigma_{id_m}\sigma_{id_{v,exp}}\rho_{id} & \sigma^2_{id_{v,exp}} \end{bmatrix}$. The m and v subscripts indicate that the effect is fitted on the mean part and the dispersion part of the model, respectively. Building on eqn 2, residuals are not assumed to be iid. $\boldsymbol{\sigma_e^2}$ is now a vector and *Diag*() is used to create a diagonal matrix. $\boldsymbol{b_v}$ and $\boldsymbol{id_v}$ are the vectors of fixed and random individual identity effects associated with the corresponding incidence matrix $\boldsymbol{X_v}$ and $\boldsymbol{Z}$. $\sigma^2_{id_{v,exp}}$ is the between-individual variance in predictability on the exponential scale. ρ_{id} is the correlation at the individual level between the variance of the mean trait and the variance of its predictability.

Finally, model 4 was a double hierarchical generalized linear animal model (DHGLAM), with a mean and a dispersion part as in model 3, but including an additive genetic component (as in model 2) fitted as a correlated random effect on both mean and predictability (Felleki *et al.*, 2012). Starting from eqn 4, model 4 can therefore be written as:

$$\begin{aligned} \boldsymbol{y}* &= \boldsymbol{X_m}\,\boldsymbol{b_m} + \boldsymbol{M}\,\boldsymbol{yr_m} + \boldsymbol{Z}\,\boldsymbol{pe_m} + \boldsymbol{Z}\,\boldsymbol{a_m} + \boldsymbol{e} \\ \log(\boldsymbol{\sigma_e^2}) &= \boldsymbol{X_v}\,\boldsymbol{b_v} + \boldsymbol{Z}\,\boldsymbol{pe_v} + \boldsymbol{Z}\,\boldsymbol{a_v} \end{aligned} \quad (5)$$

where $\begin{bmatrix} \boldsymbol{a_m} \\ \boldsymbol{a_v} \end{bmatrix} \sim N\left(\begin{matrix} 0 \\ \\ 0 \end{matrix}, \Sigma_a \otimes \boldsymbol{A}\right)$ and

$\Sigma_a = \begin{bmatrix} \sigma^2_{a_m} & \sigma_{a_m}\sigma_{a_{v,exp}}\rho_a \\ \sigma_{a_m}\sigma_{a_{v,exp}}\rho_a & \sigma^2_{a_{v,exp}} \end{bmatrix}$. $\boldsymbol{a_v}$ is the vector of additive genetic effects associated with the corresponding incidence matrix $\boldsymbol{Z}$. $\sigma^2_{a_{v,exp}}$ is the additive genetic variance of predictability on the exponential scale, and ρ_a is the additive genetic correlation between the mean and predictability variances.

The estimated variance components ($\sigma^2_{pe_{v,exp}}$ and $\sigma^2_{a_{v,exp}}$) for the predictability of docility were on the exponential scale (exp) and were converted to an additive scale ($\sigma^2_{pe_v}$ and $\sigma^2_{a_v}$) using the equations derived by Mulder *et al.* (2007) (see also Appendix S3 in Supporting Information). For each model, we estimated the phenotypic variance conditioned on the fixed effects (σ^2_P) as the sum of the variance components and the residual variance (σ^2_e). For models 3 and 4, σ^2_e was estimated as $\exp(b_{v_0}+(b_{v_1}/3)+(b_{v_2}/3))*\exp\left(\sigma^2_{yr_{v,exp}}/2\right)*\exp\left(\sigma^2_{pe_{v,exp}}/2\right)*\exp\left(\sigma^2_{a_{v,exp}}/2\right)$ (Appendix S3; Felleki *et al.*, 2012; Sae-Lim *et al.*, 2015).

Repeatability (models 1 and 3) and permanent environment effect (models 2 and 4) for mean docility was estimated as $\sigma^2_{id_m}/\sigma^2_P$ and for predictability of docility (id^2_v) as $\sigma^2_{id_v}/(2\sigma^4_P + 3(\sigma^2_{id_v} + \sigma^2_{a_v}))$. Similarly heritability of docility (h^2) and of predictability of docility (h^2_v) were estimated as $\sigma^2_{a_m}/\sigma^2_P$ and $\sigma^2_{a_v}/(2\sigma^4_P + 3(\sigma^2_{pe_v} + \sigma^2_{a_v}))$, respectively (Appendix S3; Mulder *et al.* 2007). We also estimated the genetic coefficient of variation of the variance in predictability on the additive scale as $\mathrm{GCV}_v = \sigma_{a_v}/\sigma^2_e$ Hill & Mulder, 2010).

The four models were fitted in a Bayesian framework using OpenBUGS 3.2.1 (Thomas *et al.*, 2006), run from R (R Development Core Team 2014) via the package R2OpenBUGS (Sturtz *et al.*, 2005). Quantitative genetic effects were implemented in BUGS following Waldmann (2009) and Gorjanc (2010). We used Normal priors with mean 0 and precision 0.001 for the fixed effects in both the mean and dispersion part. σ_e had a uniform prior $U(0,20)$. σ_{yr_m} and σ_{yr_v} had uniform priors $U(0,5)$. In models 1 and 2, σ_{ind_m} and σ_{pe_m} had a uniform prior $U(0,15)$, whereas in models 3 and 4, Σ_{ind} and Σ_{pe} had an inverse Wishart prior with three degrees of freedom and scale matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. In model 2, σ_{a_m} had a uniform prior $U(0,15)$, whereas for Σ_a in model 4, we used an inverse Wishart prior with three degrees of freedom and scale matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. θ_1 and θ_5 were fixed to 0 and 15, respectively, to allow identification of the parameters and facilitate convergence. Changing their values provided qualitatively similar results. The OpenBUGS code for model 4 is provided in the Supporting Information (Appendix S2). Markov chain Monte Carlo (MCMC) algorithms were iterated until convergence to the joint posterior distribution. Ten chains starting at different initial values were run in parallel. Convergence was first assessed by visually inspecting the trace plots, which were also used to identify an appropriate number of burn-in iterations. Each chain ran for 100 000 iterations including 40 000 burn-in iterations. We then checked that the Monte Carlo error was less than 1–5% of the posterior standard deviation, and that

the Brooks-Gelman-Rubin (BGR) diagnostic converged to 1 ± 0.2 (Gilks *et al.*, 1995). These convergence checks were carried out using the package coda (Plummer *et al.*, 2006) in R. The mode and 95% Highest Posterior Density Intervals (HPDI) were used to summarize the posterior distributions of the model parameters. Results are reported using the combined 600 000 iterations from the 10 unthinned chains following (Link & Eaton, 2012; Kruschke, 2014). Estimates were reported on the probit scale.

Results

We found that parameter estimates for both fixed and random effects were highly consistent across all models (Fig. 2, Table S2), validating the approach used. Between-individual variance (σ^2_{ind}) for the mean was consistent across models 1 and 3. The permanent environment (σ^2_{pe}) and additive genetic (σ^2_{pe}) variances were consistent across models 2 and 4 (Fig. 2, Table S2). In models 2 and 4, the sum of the additive genetic (σ^2_a) and permanent environment (σ^2_{pe}) variances were consistent with the estimates of σ^2_{ind} in models 1 and 3, indicating that the models were behaving adequately (Table S2). As for the mean part of the model, the sum of the additive genetic ($\sigma^2_{a_v}$) and permanent environment ($\sigma^2_{pe_v}$) variances in the dispersion part of model 4 were consistent with the estimates of $\sigma^2_{ind_v}$ in model 3 (Table S2).

Results for fixed effects in model 1 and 2 were consistent both in direction and significance with previously published studies (Petelle *et al.*, 2013, 2015). Trial number and age had a significant effect on the mean part of all models (meaning that the 95% HDPI of the associated coefficients did not include zero). Age was also significant in the dispersion part of both model 3 and 4 (Fig. 2, Table S2). However, time of the day had no significant effect when also modelling the dispersion of docility (models 3 and 4, Fig. 2, Table S2). Moreover, day of the year had a significant effect only in models 3 and 4 (Fig. 2, Table S2), when the variation in predictability was fitted in the model. In both cases, the differences in significance did not result from a change in size of the credible intervals but from a change in the mean estimates. Despite such differences, the relative coefficient estimates in both models 3 and 4 are within the credible intervals obtained from models 1 and 2 (Fig. 2, Table S2).

As previously reported (Petelle *et al.*, 2015), we found nonzero repeatability (r^2), permanent environment (pe^2) and heritability (h^2) of docility (Table 1). More importantly, we found nonzero repeatability, permanent environment and additive genetic variance in the predictability (i.e. within-individual variance) of docility (Fig. 2, Table 1). The correlation between the mean and the within-individual variance was negative at both phenotypic (cor_{ind} model 3) and permanent environment level (cor_{pe} model 4), but the associated credible intervals slightly overlapped zero (Fig. 2, Table 1). Even if these two estimates were therefore not significantly different from zero, it should be noted that 96% of their posterior distributions was negative. This suggests that cor_{ind} and cor_{pe} are negative and a larger sample size would decrease the size of the credible interval. The genetic correlation between the variance of the mean trait and the variance of predictability was not significant and was estimated as zero (Fig. 2, Table 1).

The random effect of year was negligible in the dispersion part of the models (Fig. 2, Table S2). We then refitted the models 3 and 4 without this random effect in the dispersion part of the model and found both qualitatively and quantitatively similar results.

Discussion

The long-term data available for the RMBL yellow-bellied marmot population provides a unique opportunity to advance our understanding of between-individual differences in predictability (or within-individual variance). We obtained three major results with strong implications for our understanding of behavioural evolution. First, contrary to the assumptions of most behavioural studies, we found that predictability was not homogeneous across individuals and therefore standard analytical approaches may introduce substantial bias in estimates of the mean trait. Second, we showed a significant additive genetic variance in the predictability of docility. Third, we found a negative correlation between mean docility and its predictability at the phenotypic level. This analysis is, to our knowledge, one of the first to estimate the genetic basis of both the mean of a trait and its predictability in a wild population for any trait.

Our results are quantitatively similar to previous estimates of heritability for the predictability of life-history traits, h^2_v, and their coefficient of genetic variation GCV_v obtained for captive animals (Hill & Mulder, 2010) and wild great tits (Mulder *et al.*, 2016). The estimate of h^2_v was low, partly because the predictability of docility can be affected by multiple environmental factors that reduce heritability (Houle, 1992; Westneat *et al.*, 2014). Nonetheless, its coefficient of genetic variation, GCV_v, was high (35.5%), which indicates a high potential for genetic change in response to selection (Mulder *et al.*, 2007; Hill & Mulder, 2010).

The effects of the predictability of a trait on an individual's fitness (e.g. survival or reproduction) are relatively unknown (but see Mulder *et al.*, 2015, 2016). The predictability of docility could be under selection and evolve. Different selection pressures could operate on the predictability of docility. For example, the diversity in predator species may favour individuals with higher variation in their behavioural responses. In

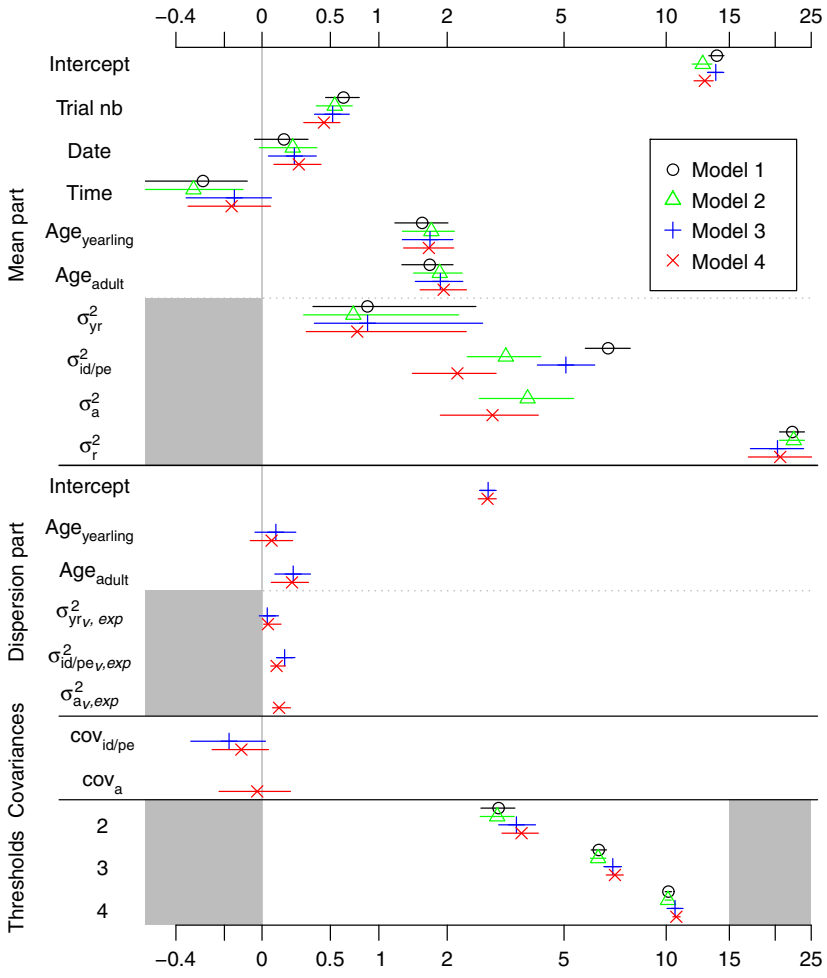

Fig. 2 Posterior mode and 95% highest posterior density intervals for the four models of docility of yellow-bellied marmots at the Rocky Mountain Biological Laboratory, Colorado. The analysis used 11 389 observations from 1576 individuals collected between 2002 and 2014. Juvenile was used as the reference level for the factor Age in both the mean and dispersion part of the model. The *x*-axis uses a logarithmic scale. Residual variances (σ^2_ε) for models 3 and 4 were estimated based on Appendix S3. σ^2 indicates variance components. The grey shaded area illustrates an invalid region of the parameter space. Variance components were constrained to be positive. Five different thresholds were needed in the ordinal model to define the six observed categories of docility. Threshold 1 and threshold 5 were fixed to 0 and 15, respectively.

addition, despite the absence of genetic correlation between docility and its predictability, predictability of docility will be under selection in the case of directional (truncation) selection on mean docility. Directional selection will not only select for the mean of the trait, but it will also select for increased trait variation (Mulder *et al.*, 2007). Individuals with lower predictability (i.e. a higher within-individual variance) have a higher probability of expressing a trait value above the selection threshold and hence be selected (Mulder *et al.*, 2007). In the marmots, selection for less docile individuals would therefore also indirectly select for lower predictability. Thus, the existence of genetic variation in predictability, despite the absence of genetic correlation with the mean, could render selection on docility less effective. The evolutionary implications of an additive genetic basis for the predictability of a trait under multivariate selection on both mean traits and their predictability are not yet fully understood (Mulder *et al.*, 2015). Correlations between the predictability of a trait and the mean level of another trait is expected to lead to a nonlinear relationship between the two traits (Mulder *et al.*, 2015). However, the impact that direct selection has on the predictability of a trait, or the consequences of a genetic correlation between the predictability of two different traits has on the evolution of mean traits, have not been investigated. As indicated by Westneat *et al.* (2014), this represents a promising area of future research.

The mechanisms underlying variance in predictability are not clear and might include phenotype switching, polyphenisms, diversification bet-hedging and multidimensional reaction norms (Westneat *et al.*, 2014). In addition to the evolutionary importance of the genetic basis of predictability in docility, we need to understand the ecological implications of the variation in predictability of docility. Optimal strategies to react to risky situations might not be only a matter of mean behaviour but also of the predictability of that behaviour. The negative phenotypic correlation between the mean docility and its predictability indicates that less docile marmots are also less predictable (i.e. lower mean

Table 1 Estimates of variance ratios of docility (with 95% highest posterior density intervals) for the four models differing in their random and residual structure. The analysis used 11 389 observations from 1576 individuals collected between 2002 and 2014.

	Models			
	1	2	3	4
σ^2_P	30.216 (27.654/33.069)	30.544 (27.925/33.322)	26.416 (22.796/31.181)	27.005 (22.704/32.106)
Mean part				
$year^2_m$	0.027 (0.011/0.08)	0.023 (0.009/0.069)	0.035 (0.014/0.096)	0.028 (0.011/0.083)
r^{2*}_m/pe^{2+}_m	0.224 (0.198/0.253)*	0.106 (0.078/0.138)$^+$	0.19 (0.158/0.225)*	0.08 (0.052/0.11)$^+$
h^2_m	–	0.126 (0.089/0.171)	–	0.106 (0.069/0.152)
Dispersion part				
$year^2_v$	–	–	0.007 (0.002/0.027)	0.009 (0.002/0.031)
r^{2*}_v/pe^{2+}_v	–	–	0.038 (0.024/0.057)*	0.024 (0.013/0.039)$^+$
h^2_v	–	–	–	0.027 (0.016/0.048)
GCV_V	–	–	–	0.355 (0.272/0.471)
Correlation				
cor^*_{ind}/cor^+_{pe}	-		−0.183 (−0.398/0.027)*	−0.262 (−0.557/0.073)$^+$
cor_a	–	–	–	−0.008 (−0.388/0.289)

σ^2_P = estimated phenotypic variance. The subscripts m and v indicate estimates for the mean or the dispersion part of the model, respectively. h^2 = heritability. r^{2*}/pe^{2+} = repeatability (models 1 and 3) or permanent environment (2 and 4) estimates. $year^2$ = proportion of variance associated with year. GCV_V is the genetic coefficient of variation for the predictability of docility. cor^*_{ind} stand for the correlation between the mean and the variance at the individual level (model 3), and cor^+_{pe} and cor_a for the permanent environment and additive genetic correlations, respectively (model 4).

docility with higher within-individual variance in docility). If docility measured in the trap reflects the antipredatory response in the wild, then, following the pace of life syndrome hypothesis (Réale *et al.*, 2010), these results could suggest that individuals that are more aggressive, explore more and thus are exposed more often to predators have a higher variance in their behavioural response. Higher variability in antipredatory behaviours for individuals exposed more often to predators could decrease predation rate and thus be an adaptive strategy. On the other hand, higher variability in antipredatory behaviours could be maladaptive because falling under a threshold of antipredatory behaviour in the presence of a predator might be deadly.

Our results showed highly consistent estimates for the random effects on mean docility across all models, whether we included heterogeneity in individual residual variance (models 3 and 4) or not (1 and 2). This suggests that ignoring the variation in predictability does not strongly bias our conclusions on the variance components of the mean docility. However, for fixed covariates, the effects of trial number and age were consistent across all models but the effects of day of the year and time of the day changed significantly between models with and without variation in predictability. Ignoring the variability in predictability (model 1 and 2) would lead to conclude that marmots are less docile in the afternoon and not affected by seasons (day of the year). However, models 3 and 4 indicated that marmots were not affected by time of the day and became more docile over the summer. Those differences could have important implications. For example, in order to handle animals when they are the most docile, the sampling protocol to do so would differ widely depending on the model used. The differences between the models for the effects of time of the day and time of the year might result from more variable individuals being sampled more often in the morning and later in the season. Overall, this result highlights that ignoring between-individual variation in predictability can bias the results obtained on the mean trait. Moreover, given the overlap in the credible intervals of these parameters across all models, it illustrates that any inference based only on a significance threshold may lead to miss important biological insights.

In this study, we introduced a framework to obtain robust estimates of the predictability of a trait that are comparable across populations, traits and studies (Appendix S3). Cleasby *et al.* (2015) proposed to estimate the coefficient of variation of predictability on the exponential scale. However, their approach only allows fitting individual identity as the sole random effect in the dispersion part of the model and was thus not appropriate in our situation. Based on work by Mulder *et al.* (2007) and Sae-Lim *et al.* (2015), we presented the equations to estimate repeatability r^2_v and heritability h^2_v of the predictability of a trait when multiple random effects are fitted in the dispersion part of the model. These equations assume that the genetic (r_a) and permanent environment (r_{pe}) correlations between the variance of the mean and of the predictability are 0 (Mulder *et al.*, 2007). Even if the consequences of this simplifying assumption seem negligible (Mulder *et al.*, 2007; Sae-Lim *et al.*, 2015), h^2_v should be used only as a first approximation in standard prediction evolutionary

models (e.g. the breeder's equation) when the genetic correlation differs from 0 (see Mulder *et al.* (2007) for the complete equations).

This is one of the first studies to have estimated the genetic variance and covariance between a trait and its predictability (i.e. within-individual variance) in a wild-living animal population. We illustrate that heritability and other variance ratios can be estimated using double hierarchical animal models and argue that these should be preferred over other techniques to compare populations, species and studies (Réale & Dingemanse, 2009; Stamps *et al.*, 2012; Cleasby *et al.*, 2015).

Acknowledgments

We thank all the 'marmoteers' who participated in collecting the long-term data and the Rocky Mountain Biological Laboratory for providing the field facilities. This work was performed using the Maxwell High Performance Computing Cluster of the University of Aberdeen IT Service (www.abdn.ac.uk/staffnet/research/research-computing), provided by Dell Inc. and supported by Alces Software. This work was funded by the UCLA Academic Senate and Division of Life Sciences, the National Geographic Society, the National Science Foundation (IDBR-0754247; DEB-1119660; DBI-0242960; DBI-07 31346), the University of Aberdeen and Marie-Curie Actions. Authors declare no conflict of interests.

Author contributions

D.T.B. led the long-term study. M.P., D.T.B. and J.G.A.M. collected data. J.G.A.M. conceived the ideas for the paper and its structure. J.G.A.M. and E.P. designed and conducted the analyses. J.G.A.M. wrote the manuscript. All authors discussed the results and edited the manuscript.

References

Armitage, K.B. 2014. *Marmot Biology. Sociality, Individual Fitness, and Population Dynamics*. Cambridge University Press, Cambridge; New York.

Blumstein, D.T., Wey, T.W. & Tang, K. 2009. A test of the social cohesion hypothesis: interactive female marmots remain at home. *Proc. Roy. Soc. B* **276**: 3007–3012.

Blumstein, D.D., Lea, A.A., Olson, L.L. & Martin, J.G.A.J. 2010. Heritability of anti-predatory traits: vigilance and locomotor performance in marmots. *J. Evol. Biol.* **23**: 879–887.

Byers, B.E. & Kroodsma, D.E. 2009. Female mate choice and songbird song repertoires. *Anim. Behav.* **77**: 13–22.

Cleasby, I.R., Nakagawa, S. & Schielzeth, H. 2015. Quantifying the predictability of behaviour: statistical approaches for the study of between-individual variation in the within-individual variance. *Methods Ecol. Evol.* **6**: 27–37.

Dingemanse, N.J. & Dochtermann, N.A. 2013. Quantifying individual variation in behaviour: mixed-effect modelling approaches. *J. Anim. Ecol.* **82**: 39–54.

Felleki, M., Lee, D., Lee, Y., Gilmour, A.R. & Rönnegård, L. 2012. Estimation of breeding values for mean and dispersion, their variance and correlation using double hierarchical generalized linear models. *Genet. Res.* **94**: 307–317.

Foulley, J.L. & Jaffrézic, F. 2010. Modelling and estimating heterogeneous variances in threshold models for ordinal discrete data via Winbugs/Openbugs. *Computer Methods Programs Biomed.* **97**: 19–27.

Frase, B. & Hoffmann, R. 1980. *Marmota flaviventris*. *Mammal. Sp.* **135**: 1–8.

Gilks, W.R., Richardson, S. & Spiegelhalter, D. 1995. *Markov Chain Monte Carlo in Practice*. CRC Press, Boca Raton, FL.

Gorjanc, G. 2010. Flexible Bayesian inference of animal model parameters using BUGS program. In: *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production*. Leipzig, Germany.

Hill, W.G. & Mulder, H.A. 2010. Genetic analysis of environmental variation. *Genet. Res.* **92**: 381–395.

Houle, D. 1992. Comparing evolvability and variability of quantitative traits. *Genetics* **130**: 195–204.

Kalinowski, S.T., Taper, M.L. & Marshall, T.C. 2007. Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* **16**: 1099–1106.

Kruschke, J. 2014. *Doing Bayesian Data Analysis: A Tutorial With R, JAGS, and Stan*. Academic Press. https://books.google.co.uk/books?id=FzvLAwAAQBAJ.

Kruuk, L.E.B. 2004. Estimating genetic parameters in natural populations using the 'animal model'. *Phil. Trans. Roy. Soc. B* **359**: 873–890.

Lee, Y. & Nelder, J.A. 2006. Double hierarchical generalized linear models – discussion. *J. Roy. Stat. Soc. C* **55**: 139–185.

Link, W.A. & Eaton, M.J. 2012. On thinning of chains in MCMC. *Methods Ecol. Evol.* **3**: 112–115.

Mulder, H.A., Bijma, P. & Hill, W.G. 2007. Prediction of breeding values and selection responses with genetic heterogeneity of environmental variance. *Genetics* **175**: 1895–1910.

Mulder, H.A., Hill, W.G. & Knol, E.F. 2015. Heritable environmental variance causes nonlinear relationships between traits: application to birth weight and stillbirth of pigs. *Genetics* **199**: 1255–1269.

Mulder, H.A., Gienapp, P. & Visser, M.E. 2016. Genetic variation in variability: phenotypic variability of fledging weight and its evolution in a songbird population. *Evolution* **70**: 2004–2016.

Nussey, D.H., Wilson, A.J. & Brommer, J.E. 2007. The evolutionary ecology of individual phenotypic plasticity in wild populations. *J. Evol. Biol.* **20**: 831–844.

Olson, L.E., Blumstein, D.T., Pollinger, J.R. & Wayne, R.K. 2012. No evidence of inbreeding avoidance despite demonstrated survival costs in a polygynous rodent. *Mol. Ecol.* **21**: 562–571.

Petelle, M.B., McCoy, D.E., Alejandro, V., Martin, J.G.A. & Blumstein, D.T. 2013. Development of boldness and docility in yellow-bellied marmots. *Anim. Behav.* **86**: 1147–1154.

Petelle, M.B., Martin, J.G.A. & Blumstein, D.T. 2015. Heritability and genetic correlations of personality traits in a wild population of yellow-bellied marmots (*Marmota flaviventris*). *J. Evol. Biol.* **28**: 1840–1848.

Pigliucci, M. 2005. Evolution of phenotypic plasticity: Where are we going now?. *Trends Ecol. Evol.* **20**: 481–486.

Pinheiro, J.C. & Bates, D.M. 2000. *Mixed-Effects Models in S and S-PLUS*. Springer Verlag, New York.
Plummer, M., Best, N., Cowles, K. & Vines, K. 2006. CODA: convergence diagnosis and output analysis for MCMC. *R News* **6**: 7–11.
R Development Core Team. 2014. *R: A Language and Environment for Statistical Computing*. R Development Core Team, Vienna, Austria.
Ram, N. & Gerstorf, D. 2009. Time-structured and net intraindividual variability: tools for examining the development of dynamic characteristics and processes. *Psychol. Aging* **24**: 778–791.
Réale, D. & Dingemanse, N.J. 2009. Personality and individual social specialisation. In: *Social Behaviour: Genes, Ecology and Evolution* (T. Szekely, A.J. Moore & J. Komdeur, eds). Cambridge University Press, Cambridge, UK.
Réale, D., Reader, S.M., Sol, D., McDougall, P.T. & Dingemanse, N.J. 2007. Integrating animal temperament within ecology and evolution. *Biol. Rev.* **82**: 291–318.
Réale, D., Garant, D., Humphries, M.M., Bergeron, P., Careau, V. & Montiglio, P.O. 2010. Personality and the emergence of the pace-of-life syndrome concept at the population level. *Phil. Trans. Roy. Soc. B* **365**: 4051–4063.
Roff, D.A. 2002. *Life History Evolution*. Sinauer Associates, Sunderland, MA.
Rönnegård, L., Felleki, M., Fikse, F., Mulder, H.A. & Strandberg, E. 2010. Genetic heterogeneity of residual variance – estimation of variance components using double hierarchical generalized linear models. *Genet. Sel. Evol.* **42**: 8.
Sae-Lim, P., Kause, A., Janhunen, M., Vehviläinen, H., Koskinen, H., Gjerde, B. *et al.* 2015. Genetic (co)variance of rainbow trout (*Oncorhynchus mykiss*) body weight and its uniformity across production environments. *Genet. Sel. Evol.* **47**: 46.
SanCristobal-Gaudy, M., Elsen, J.M., Bodin, L. & Chevalet, C. 1998. Prediction of the response to a selection for canalisation of a continuous trait in animal breeding. *Genet. Sel. Evol.* **30**: 423–451.
Stamps, J.A., Briffa, M. & Biro, P.A. 2012. Unpredictable animals: individual differences in intraindividual variability (IIV). *Anim. Behav.* **83**: 1325–1334.
Sturtz, S., Ligges, U. & Gelman, A. 2005. R2WinBUGS: a package for running WinBUGS from R. *J. Stat. Soft.* **12**: 1–16.
Thomas, A., O'Hara, B., Ligges, U. & Sturtz, S. 2006. Making BUGS open. *R News* **6**: 12–17.
Waldmann, P. 2009. Easy and flexible Bayesian inference of quantitative genetic parameters. *Evolution* **63**: 1640–1643.
Westneat, D.F., Schofield, M. & Wright, J. 2013. Parental behavior exhibits among-individual variance, plasticity, and heterogeneous residual variance. *Behav. Ecol.* **24**: 598–604.
Westneat, D.F., Wright, J. & Dingemanse, N.J. 2014. The biology hidden inside residual within-individual phenotypic variation. *Biol. Rev.* **90**: 729–743.
Wilson, A.J., Réale, D., Clements, M.N., Morrissey, M.M., Postma, E., Walling, C.A. *et al.* 2009. An ecologist's guide to the animal model. *J. Anim. Ecol.* **79**: 13–26.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1 Supplementary tables and figures.
Appendix S2 Annotated code to fit model 4 in OpenBugs (File: MartinSA2.txt).
Appendix S3 Calculation of variance components and variance ratios for (exponential) double hierarchical models.
Table S1 Summary statistics for the pruned pedigree used for animal models of docility of yellow-bellied marmots at RMBL.
Table S2 Parameter estimates (with 95% highest posterior density intervals) from four models of docility of yellow-bellied marmots.
Figure S1 Mean (point) and range (dotted line) of docility for 1576 yellow-bellied marmots at RMBL between 2002–2014. Individuals were sorted by mean docility.

Data deposited at Dryad: doi:10.5061/dryad.11vf0.

Received 5 April 2016; revised 1 February 2017; accepted 2 February 2017