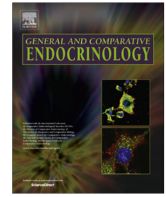
General and Comparative Endocrinology 208 (2014) 154–163

Contents lists available at ScienceDirect

General and Comparative Endocrinology

journal homepage: www.elsevier.com/locate/ygcen



Baseline and stress-induced glucocorticoid concentrations are not repeatable but covary within individual great tits (*Parus major*)

Alexander T. Baugh [a,b,*], Kees van Oers [c], Niels J. Dingemanse [d,e], Michaela Hau [b,f,g]

[a] Department of Biology, Swarthmore College, 500 College Avenue, Swarthmore, PA 19081, USA
[b] Evolutionary Physiology Group, Max Planck Institute for Ornithology, Am Obstberg 1, 78315 Radolfzell, Germany
[c] Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), P.O. Box 50, 6708 PB Wageningen, The Netherlands
[d] Evolutionary Ecology of Variation Research Group, Max Planck Institute for Ornithology, Eberhard-Gwinner-Str., 82319 Seewiesen, Germany
[e] Behavioural Ecology, Department of Biology, Ludwig-Maximilians University of Munich, Planegg-Martinsried, Germany
[f] Evolutionary Physiology Group, Max Planck Institute for Ornithology, Eberhard-Gwinner-Str., 82319 Seewiesen, Germany
[g] Department of Biology, University of Konstanz, Universitätsstrasse 10, Konstanz, Germany

ARTICLE INFO

Article history:
Received 15 April 2014
Revised 19 August 2014
Accepted 22 August 2014
Available online 10 September 2014

Keywords:
Corticosterone
Endocrine stress response
Hypothalamic-pituitary-adrenal axis
Multi-level variation
Repeatability
Variance partitioning

ABSTRACT

In evolutionary endocrinology, there is a growing interest in the extent and basis of individual variation in endocrine traits, especially circulating concentrations of hormones. This is important because if targeted by selection, such individual differences present the opportunity for an evolutionary response to selection. It is therefore necessary to examine whether hormone traits are repeatable in natural populations. However, research in this area is complicated by the fact that different hormone traits can be correlated. The nature of these trait correlations (i.e., phenotypic, within-, or among-individual) is critically relevant in terms of the evolutionary implications, and these in turn, depend on the repeatability of each hormone trait. By decomposing phenotypic correlations between hormone traits into their within- and among-individual components it is possible to describe the multivariate nature of endocrine traits and generate inferences about their evolution. In the present study, we repeatedly captured individual great tits (*Parus major*) from a wild population and measured plasma concentrations of corticosterone. Using a mixed-modeling approach, we estimated repeatabilities in both initial (cf. baseline; CORT0) and stress-induced concentrations (CORT30) and the correlations between those traits among- and within-individuals. We found a lack of repeatability in both CORT0 and CORT30. Moreover, we found a strong phenotypic correlation between CORT0 and CORT30, and due to the lack of repeatability for both traits, there was no among-individual correlation between these two traits—i.e., an individual's average concentration of CORT0 was not correlated with its average concentration of CORT30. Instead, the phenotypic correlation was the result of a strong within-individual correlation, which implies that an underlying environmental factor co-modulates changes in initial and stress-induced concentrations within the same individual over time. These results demonstrate that (i) a phenotypic correlation between two hormone traits does not imply that the traits are correlated among individuals; (ii) the importance of repeated sampling to partition within- and among-individual variances and correlations among labile physiological traits; and (iii) that environmental factors explain a considerable fraction of the variation and co-variation in hormone concentrations.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

1.1. Theoretical background: levels of phenotypic variation

Evolution requires heritable variation, but estimating the heritability of traits in free-living animals is often impractical due to the absence of pedigree information. Under these conditions it is often useful to estimate the repeatability of traits, which sets an upper limit to heritability (Lessells and Boag, 1987; but see Dohm, 2002). Evidence for repeatable variation therefore provides clues as to whether the observed trait can in principle evolve in response to selection (e.g., Dingemanse and Dochtermann, 2014).

* Corresponding author at: Swarthmore College, 500 College Avenue, Martin Biological Laboratory, Swarthmore PA 19081, USA. Fax: +1 610 328 8663.
E-mail addresses: abaugh1@swarthmore.edu (A.T. Baugh), k.vanoers@nioo.knaw.nl (K. van Oers), ndingemanse@orn.mpg.de (N.J. Dingemanse), mhau@orn.mpg.de (M. Hau).

http://dx.doi.org/10.1016/j.ygcen.2014.08.014
0016-6480/© 2014 Elsevier Inc. All rights reserved.

Morphological measures such as wing length or body size are examples of traits that typically have high repeatability. Repeatability is the proportion of phenotypic variance explained by among-individual variance—thus, traits with low within- and high among-individual variance have high repeatability. Within-individual variance describes the amount of phenotypic variation among observations of the same individual over time (Fig. 1) and therefore represents the plasticity (as well as stochasticity and measurement error) of an individual's phenotype in response to external (e.g., ambient temperature) and internal variables (e.g., nutritional status, age). Within-individual variation is therefore not synonymous with plasticity, which instead represents only the portion of this within-individual variation that is due to the individual's response to environmental change.

Among-individual variances describe how much individuals differ from each other in their average phenotype (Fig. 1). Among-individual variation, therefore, represents the more static aspects of the phenotype, which can arise due to 'permanent' external factors (e.g., maternal effects or environmental influences that are stable over the course of the measurements) or heritable genetic differences—fulfilling, in the case of the latter, a prerequisite for evolutionary responses to selection. Within- and among-individual variances thus represent two hierarchical levels of variation and jointly contribute to the phenotypic (i.e., total) variance (Westneat et al., 2014; Fig. 1).

Using heritability estimates (h^2), it is then possible to calculate the evolutionary response (R) to selection (S) using the classic breeder's equation ($R = h^2 * S$). Evolution, however, will proceed very differently if the focal trait is genetically correlated with other traits under selection (Lande and Arnold, 1983). It is therefore imperative that we quantify correlations between traits in order to be able to properly predict how they might evolve. The most pressing consequence of genetic correlations is that they can impose constraints on evolution; the constraining effects of these correlations can be substantial for behavioral (Dingemanse and Dochtermann, 2013) and endocrine traits (Ketterson et al., 2009).

Phenotypic correlations between two traits are sometimes used to infer their genetic correlation, which might be a suitable assumption for certain classes of traits, particularly non-labile (i.e., fixed) traits such as morphological dimensions (cf. Cheverud's conjecture; Dochtermann, 2011). This inference, however, is more complicated for labile (i.e., plastically varying) traits. Raw phenotypic correlations for labile traits are the weighted outcome of two processes: (i) 'within-individual correlations' which are the integration of plasticity between two focal traits (cf. both traits change in concert within the same individual due to one or more environmental variables); and (ii) 'among-individual correlations' which are proximately underpinned by the effects of maternal, 'permanent' environmental and genetic correlations between the traits, with genetic correlations occurring either because of

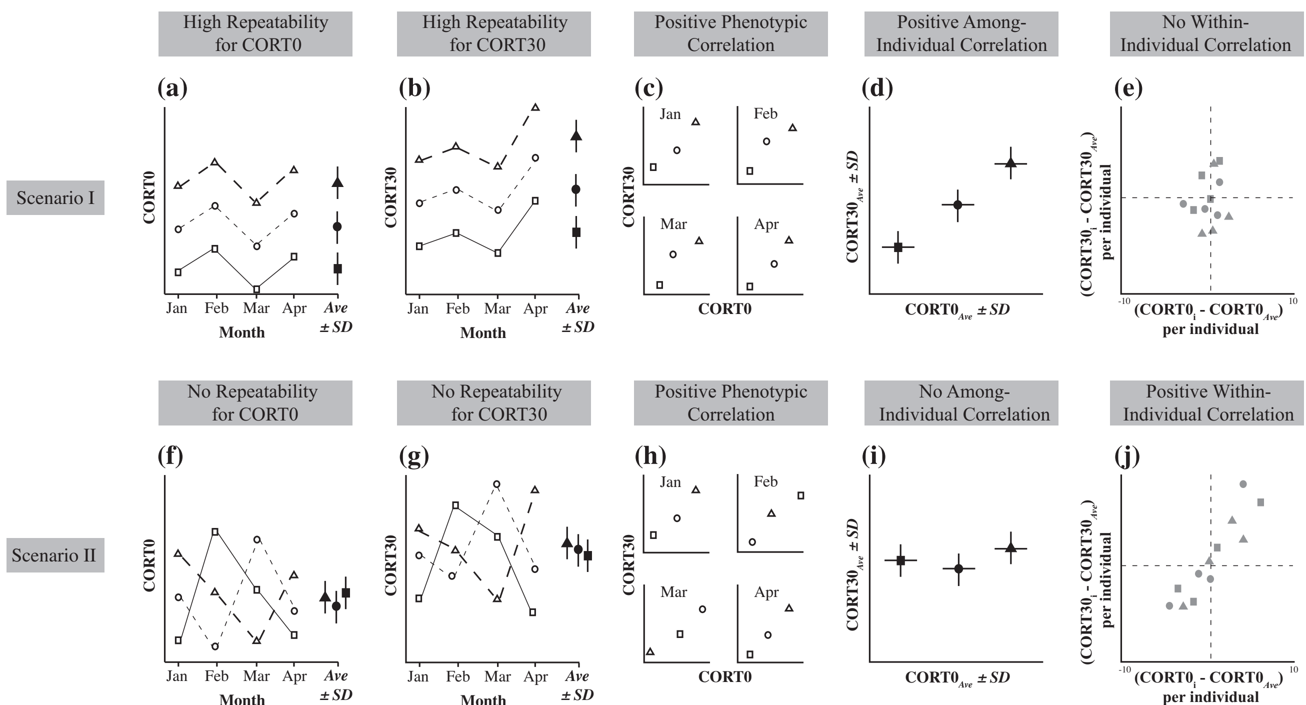


Fig. 1. Illustrations of two hypothetical scenarios in which concentrations of CORT0 and CORT30 are measured at four time points in each of three individuals (square, circle, triangle). Scenario I depicts a situation in which the three individuals differ consistently from each other in both CORT0 and CORT30, as indicated by the consistent rank order of each subject (e.g., triangles are always above circles at any given month) in (a) and (b). Therefore, averages for both CORT0 and CORT30 in these three individuals are distinct (solid symbols in (a) and (b)) and correspondingly, repeatability (i.e., among individual variance) is high. In any given month, the rank order of individuals for CORT0 is the same as that for CORT30, leading to a positive phenotypic correlation between these two hormone concentrations (c). This positive phenotypic correlation is driven by a positive among-individual correlation, and can be illustrated by showing the individual averages for CORT0 versus CORT30 (d), which is due to the fact that the rank order of individuals is stable across months (e.g., triangles are always highest for both CORT0 and CORT30). Note that error bars in (d) illustrate that there is also some within-individual variance in both CORT0 and CORT30. Within-individual correlations can be depicted by plotting the deviation from the average per individual for each measure (i) of CORT0 (x-axis) versus the deviation from the average per individual for each measure (i) of CORT30 (y-axis). The lack of a within-individual correlation is depicted in (e), indicating that the phenotypic correlations in (c) are driven principally by the among-individual correlation (d). Scenario II depicts an alternative situation in which each individual varies considerably from one month to the next leading to a lack of repeatability (i.e., no among-individual variance) for both CORT0 (f) and CORT30 (g). Nevertheless, in any given month, the ranking of individuals for CORT0 is the same as that for CORT30, leading to a phenotypic correlation (h) similar to Scenario I. Because of the lack of repeatability, this phenotypic correlation cannot be driven by an among-individual correlation, which is depicted by the lack of relationship in the average CORT0 and CORT30 phenotypes (i). Instead, it must be driven by a within-individual correlation (j), indicating the role of environmental factors in co-modulating CORT0 and CORT30 concentrations simultaneously within the individual. Scenario II better illustrates the results from the present study.

pleiotropic effects or linkage disequilibrium (Roff, 1997). Raw phenotypic correlations can have non-zero values both in the presence and absence of within- or among-individual variation (Eq. (1); Fig. 1).

If among- and within-individual correlations are the same, the raw phenotypic correlation accurately approximates the among-individual correlation (Dingemanse and Dochtermann, 2013); however, if they are different in sign, which is not uncommon—because those correlations are the outcome of very different processes—then this is not the case. Hence, it is important to partition the raw phenotypic correlation into its among- and within-individual components. In Fig. 1 we illustrate two hypothetical scenarios that represent the ends of a continuum for how variances in single endocrine traits and correlations between them can be manifested at the phenotypic, within- and among-individual levels.

1.2. Empirical study: levels of variance and correlation in glucocorticoid concentrations in wild birds

In comparative endocrinology it is common to measure hormone concentrations in individual animals only once and assume that the variation observed among those individuals represents among-individual variance (for a discussion of this topic, see Williams, 2008). However, hormones act as mediators between the internal (e.g., glucose concentrations) and external environments (e.g., ambient temperature), and therefore their concentrations often vary from moment-to-moment. Therefore, repeated measures designs are necessary to quantify among-individual, i.e., repeatable, variation (Williams, 2008). The often mixed results on the topic of hormonal repeatability (Cockrem and Silverin, 2002b; Duckworth and Sockman, 2012; Jawor et al., 2006; Kralj-Fiser et al., 2007; Ouyang et al., 2011; Patterson et al., 2014; Rensel and Schoech, 2011; Romero and Reed, 2008; Vitousek et al., 2008; Wada et al., 2008; While et al., 2010), might be due in part to the fact that while some studies have based their inferences about individual-level processes on repeated measures data, others have not. Studies not using repeated measures designs and variance partitioning might be reporting unrepeatable (cf. within-individual) parts of the phenotype instead (Dingemanse et al., 2012a; Dingemanse and Dochtermann, 2013; Williams, 2008).

With a growing interest in the role of hormones in evolution (Hau, 2007; Heideman, 2004; Ketterson et al., 1996; McGlothlin and Ketterson, 2008; McGlothlin et al., 2010; Zera et al., 2007), it is critical that we know whether hormonal traits are (i) repeatable and (ii) correlated at the among- or the within-individual level, or both. The hormonal response to stress presents an important system to understand in this way because of its activational role in responding to environmental challenges and the potential correlations among its multiple components (Bókony et al., 2009; Eikenaar et al., 2012; Goymann et al., 2004; Hau et al., 2010; Koolhaas et al., 1999; Romero, 2004).

In vertebrates, the hypothalamic–pituitary–adrenal (HPA) axis performs two fundamental tasks: (1) maintaining homeostasis during everyday life and (2) coordinating the stress response. Under typical conditions the HPA axis regulates baseline concentrations of circulating glucocorticoids, which help coordinate metabolic demands in response to predictable (periodic) environmental fluctuations, such as day-night rhythmicity in metabolism and activity (Landys et al., 2006). The HPA axis fulfills another role by responding to unpredictable, often 'stressful' events, such as agonistic interactions with conspecifics (Carere et al., 2003; Landys et al., 2010; Van Duyse et al., 2004); but see (Wingfield and Lewis, 1993), inclement weather (Breuner and Hahn, 2003), or predators (Cockrem and Silverin, 2002a; Eilam et al., 1999; but see Butler et al., 2009). This stress response is initiated within a few minutes after stressor onset (e.g., capture), and results in secretion of glucocorticoids above baseline concentrations. Glucocorticoid concentrations then continue to increase in the blood until they reach a peak concentration, typically within 30–90 min in avian capture-handling-restraint studies (Baugh et al., 2013; Wingfield et al., 1982). These elevated concentrations of circulating glucocorticoids redistribute glucose stores to critical tissues (e.g., skeletal muscle) to support this 'emergency life history stage' (Wingfield et al., 1998). A process of negative feedback subsequently reduces circulating glucocorticoid concentrations, thereby enabling baseline concentrations to be re-achieved. Corticosterone (CORT) is the main glucocorticoid in birds, and like many steroid hormones, CORT can affect diverse regulatory and behavioral processes simultaneously (cf. hormonal pleiotropy; Almasi et al., 2013; Baugh et al., 2012, 2013; Carere et al., 2010; Harri et al., 2003; McGlothlin and Ketterson, 2008; van Oers et al., 2011). Importantly, these different components of glucocorticoid physiology are known to be interdependent (Romero, 2004), but are often analyzed as though they are independent dimensions.

Given this interdependence, it is important to address whether these measures are correlated at the phenotypic, among- and within-individual levels. Among-individual correlations, for example, in initial CORT (hereafter CORT0) versus stress-induced CORT (hereafter CORT30), can only exist when both of these traits show among-individual variation (Dingemanse and Dochtermann, 2013; Fig. 1). An among-individual correlation would exist if the average CORT0 phenotype were correlated with the average CORT30 phenotype from the same set of individuals (Fig. 1).

Similarly, within-individual correlations can only be present when both traits show within-individual variation; significant within-individual correlations would exist when the change in CORT0 across observations correlates with the change in CORT30 across those same observations within the same individual. This would occur, for example, when (i) the expression of both traits is underpinned by a common factor and this common factor varies within an individual over time, such as nutritional status, workload, or environmental conditions; and (ii) when each of the two traits varies in response to a different environmental factor but those two factors (e.g., conspecific density and predation risk) are themselves correlated due to another process (e.g., habitat quality). Importantly, studies that do not include repeated measurements are limited to estimates of correlations that are not partitioned, which means that distinguishing between acute (e.g., environmental) versus permanent (e.g., genetic) sources of variation in traits and correlations between them is not possible.

In the present study, we used a population of wild great tits (*Parus major*) and mixed effects modeling to address the following questions: (1) Are initial and stress-induced concentrations of CORT individually repeatable—a necessary pre-condition for these traits to evolve in response to selection? (2) Are initial and stress-induced concentrations correlated at the phenotypic level and if so, is this phenotypic correlation driven by a within- or among-individual correlation or both? (3) What does such a correlation indicate about the mechanistic link between these two traits? We believe the statistical approaches used here, which are quickly becoming mainstream in fields such as behavioral ecology, hold promise for better elucidating the causes and potential evolutionary consequences of individual variation in endocrine phenotypes.

2. Materials and methods

2.1. Sample collection

In the autumn of 2010 we sampled birds from a long-term study population of ringed wild great tits (*P. major*) in the Westerheide study area near Arnhem, The Netherlands (52° 0′ 38″ N, 5°

50′ 30″ E, 35 m AMSL). We captured adult birds for the measurement of plasma CORT using a standardized capture-handling-restraint protocol to examine baseline and stress-induced concentrations (Romero et al., 1997). We conducted this study in autumn to minimize any potential influence of breeding or molt, thereby restricting our study to a single life history stage and minimizing variation arising due to seasonal influences. Further, by sampling during a restricted time of day (0800–1100), we aimed to minimize variation in hormone concentrations resulting from diel factors (Breuner et al., 1999).

We suspended mist nets adjacent to five feeding stations and observed from a distance of ca. 20 m behind vegetation. The moment a bird intercepted the net, we initiated a digital timer, rapidly removed the bird from the net and collected approximately 40 μL of blood at each of two time points by puncturing the brachial vein using a 30-gauge needle. All CORT0 blood samples were collected in less than 3 min following entry into the net—most passerine species exhibit a detectable increase in plasma CORT approximately 3 min after the onset of an acute stressor (Baugh et al., 2013; Romero and Reed, 2005; Romero and Romero, 2002). We refer to the duration of time between entry into the net and the completion of the CORT0 bleed as 'CORT0 handling time' and we refer to this concentration as 'initial' instead of 'baseline' because we define 'baseline' as the average concentration in an unstressed individual correcting for diel modulation—a trait difficult to measure in any organism, but particularly in free-living animals. After the initial bleed, birds were held in cloth bags until 30 min after capture, when they were re-bled (CORT30). The duration of time between opening the cloth bag and completing the CORT30 bleed is referred to as 'CORT30 handling time.' We chose the 30 min time point in an attempt to capture robustly elevated concentrations without the need for protracted periods of restraint. Previous research in this population showed that glucocorticoid concentrations are universally elevated at 30 min post-capture; however, there is variation in the temporal profiles with some birds reaching maximal concentrations before or after this time point (Baugh et al., 2013). We therefore do not assume that the CORT30 values represent maximal concentrations, but rather a single measure of the strength of the stress response under standardized conditions.

After the 30 min sample, we fitted unbanded birds with a uniquely numbered aluminum ring (Vogeltrek Station, the Netherlands) and immediately released them at the site of capture. Fifty-eight adult birds were recaptured 2–4 times each (2 repeats for 44 birds; 3 repeats for 12 birds; 4 repeats for 2 birds; 132 captures; 264 blood samples) with a median interval across captures of 11 days (mean ± SD: 13.5 ± 11.9 d; range: 3–82 d). We tested the possibility that the interval (days) across captures influences CORT concentrations (Lynn et al., 2010). To maintain consistency in capture, handling and bleeding practices, all sampling was conducted by a single person with considerable prior experience (ATB). We maintained the blood samples on wet ice for less than 3 h prior to centrifugation (5000 rpm for 10 min; ca. 1400*g*) and stored the plasma fraction at −80 °C until November 2010 when samples were transferred on dry ice to the Max Planck Institute for Ornithology (Radolfzell, Germany) for hormone measurement.

2.2. Enzyme immunoassay for corticosterone

We used standard enzyme immunoassay techniques to estimate plasma CORT concentrations (Enzo Life Sciences, Cat. No. ADI 900-097; Donkey anti-Sheep IgG). The details of our EIA protocol validation have been reported elsewhere (Ouyang et al., 2011). Briefly, concentrations were determined following a diethyl-ether extraction of a 10 μL sample volume. After drying extracts under a stream of N_2 gas, samples were diluted (1:30) using Tris-buffered saline (supplied by kit) and samples were allowed to reconstitute overnight at 4 °C. Samples were then assayed in duplicate along with blanks, five standards (0.032–20 ng mL^{-1} CORT) and positive controls. Final values were corrected for average recovery loss, which we determined previously using individual samples spiked with radioactively labeled CORT (mean ± SD; ca. 85% ± 2.7). The intra- and inter-assay coefficients of variation (CV)—7.79% and 8.32%, respectively—were determined by distributing a minimum of 2 duplicate samples of stripped chicken plasma spiked with commercial CORT (supplied by kit) to a concentration of 20 ng mL^{-1} across each of the 11 plates. Additionally, to estimate the 'technical repeatability' of the immunoassay, we reprocessed a subset of experimental plasma samples ($n = 23$). This "technical repeatability" for our procedure was very high ($r = 0.962 \pm 0.015$ (SE); p 0.0001; $n = 23$ birds sampled twice each). Note that this includes non-assay sources of variation (extraction, pipetting, freeze–thaw cycles), thus providing a cumulative estimate of measurement error in plasma CORT concentrations. The assay has a detection limit of 27 pg mL^{-1}. The cross-reactivity of the antiserum is 100% for corticosterone, 28.6% for deoxycorticosterone and 1.7% for progesterone.

2.3. Statistical analyses

We considered initial and stress-induced concentrations of CORT as separate traits in our models—an approach consistent with an established assumption that these two aspects are differentially regulated by subsystems of the HPA axis (mineralocorticoid and glucocorticoid receptors in the brain), and that variation in these two measures have different consequences for the organism (Landys et al., 2006; Romero, 2004; Schmidt et al., 2012). Therefore, we used a 'character state' approach to analyze the sources of variation in the data (Via et al., 1995). This is similar to a 'reaction norm' approach for our particular study since these two approaches converge when modeling a two response variable system (Roff, 1997), with the exception that the character state approach does not assume that the residual error variance is equal for the two traits—this variance component is instead simply estimated directly. It is important to note that although we model CORT0 and CORT30 as separate traits, we examine this assumption by evaluating the within-versus the among-individual correlations for these two hormone traits.

All hormone data were $\log_{10}$ transformed to satisfy assumptions of normality and centered by conversion to standardized (z) scores. We used a two-step statistical approach to estimate within- and among-individual variance in CORT0 and CORT30 and the within- and among-individual correlations between these two traits. Note, that our use of 'correlation' between two traits is not to be confused with correlational approaches used to estimate repeatability of single traits, such as intra-class correlations (Lessells and Boag, 1987). First, we fitted univariate mixed-effects models with a Gaussian error distribution for each trait separately, for which random intercepts for individual identity were fitted to estimate the among- and within-individual variance as well as repeatability. Second, we fitted a bivariate mixed-effects model with random intercepts for individual identity and with CORT0 and CORT30 as the two response variables, assuming a bivariate Gaussian distribution. This bivariate model allowed us to estimate the phenotypic correlation between the two traits $\boldsymbol{r}_{p_y,p_z}$ and decompose it into its among- and within-individual components, as detailed in Eq.(1) (Dingemanse and Dochtermann, 2013):

$$\begin{aligned} \boldsymbol{r}_{p_y,p_z} = \boldsymbol{r}_{ind_{0y},ind_{0z}} \\ \times \sqrt{\left(\boldsymbol{v}_{ind_{0y}} \div \left(\boldsymbol{v}_{ind_{0y}} + \boldsymbol{v}_{e_{0y}}\right)\right) \times \left(\boldsymbol{v}_{ind_{0z}} \div \left(\boldsymbol{v}_{ind_{0z}} + \boldsymbol{v}_{e_{0z}}\right)\right)} \\ + \boldsymbol{r}_{e_{0y},e_{0z}} \\ \times \sqrt{\left(\boldsymbol{v}_{e_{0y}} \div \left(\boldsymbol{v}_{ind_{0y}} + \boldsymbol{v}_{e_{0y}}\right)\right) \times \left(\boldsymbol{v}_{e_{0z}} \div \left(\boldsymbol{v}_{ind_{0z}} + \boldsymbol{v}_{e_{0z}}\right)\right)} \end{aligned} \quad (1)$$

where $\boldsymbol{r}_{p_y,p_z}$, $\boldsymbol{r}_{ind_{0y},ind_{0z}}$, and $\boldsymbol{r}_{e_{0y},e_{0z}}$ represent the phenotypic, among-individual and within-individual correlations, respectively; $\boldsymbol{v}_{ind_{0y}}$ and $\boldsymbol{v}_{ind_{0z}}$ are the among-individual variances; and $\boldsymbol{v}_{e_{0y}}$ and $\boldsymbol{v}_{e_{0z}}$ represent the within-individual variances; with subscripts y and z representing CORT0 and CORT30, respectively.

The statistical significance of random effects was determined by a likelihood ratio test. The test statistic is twice the difference in log-likelihood between hierarchical models, and is distributed as χ^2 with degrees of freedom equal to the difference in the number of (co)variance parameters estimated (Meyer, 1992).

Because of three statistical outliers in the data (CORT0 samples from 3 birds), we constructed a variable called 'outlier' (yes/no) and ran our mixed-effects models both with and without fitting this variable as an additional fixed effect. We flagged these three CORT0 samples along with their associated CORT30 samples because they greatly exceeded the average CORT0 concentrations (>3 SD above mean) and therefore likely represent animals that had been recently stressed prior to capture. This approach enabled us to determine whether any of the observed patterns in the data were caused by these statistical outliers while ensuring that these models did not differ in sample size. Mixed-effects models were fitted using the program ASReml v3.0 (VSN International).

3. Results

3.1. General

Our handling-restraint stress protocol resulted in a universal increase in circulating CORT between the 0 and 30 min time points (mean ± SEM: CORT0: 5.02 ng mL^{-1} ± 0.35; CORT30: 20.20 ng mL^{-1} ± 0.98; paired t-test: $P < 0.0001$). The average CORT0 handling time was 135.9 s (SD = 32 s; range = 50–179 s) and the average CORT30 handling time was 120.9 s (SD = 70 s; range = 2–480 s). We investigated the influence of a set of potentially confounding covariates (handling times; blood volumes; date of capture; days between repeated captures; body condition) and factors (sex; sampling repeat number within each bird—i.e., first, second, third, or fourth capture). These covariates and factors, however, did not change the statistical significance of the results (see Supplemental Materials).

3.2. Repeatability and levels of variance

Our models indicated that CORT0 and CORT30 were not significantly repeatable (r = 0.13–0.26, all $P > 0.05$) (Table 1). Within-individual variance always exceeded among individual variance. As a representative example, this lack of individual repeatability is illustrated in a plot of concentrations at first versus second capture (Fig. 2). This finding implies either that those traits were not repeatable or that non-zero repeatability was too low to detect statistically. The inclusion versus exclusion of outliers in the dataset did not influence the statistical significance of the results (Table 1).

3.3. Levels of correlation

There was a strong and positive phenotypic correlation between CORT0 and CORT30 (Fig. 3a–c; Table 2). This means that captures in which initial concentrations were high also had high

Table 1
Univariate mixed-effects models showing the within- and among-individual variance components and a significance test of the repeatability estimate. Statistical outliers (n = 3) were included in all models and were either fitted as a fixed effect (*) or not.

Fixed Effects						Variance Components						Repeatability			
Model	Trait	Intercept	SE	Outlier	SE	Total	SE	Among	SE	Within	SE	r	SE	Chi^2	P
1	Log_{10}CORT0	0.576	0.01926	NA	NA	0.084101	0.007607	0.022216	0.010946	0.061885	0.010432	0.264	0.122	2.312	0.128
1*	Log_{10}CORT0	0.567	0.01770	0.7774	0.1409	0.074392	0.006611	0.009885	0.010745	0.064507	0.011260	0.133	0.142	0.238	0.626
2	Log_{10}CORT30	1.190	0.01798	NA	NA	0.074455	0.006672	0.014675	0.008969	0.059780	0.009335	0.197	0.116	2.414	0.120
2*	Log_{10}CORT30	1.183	0.01808	0.406	0.1406	0.072893	0.006596	0.017538	0.008923	0.055356	0.008851	0.241	0.116	3.402	0.065

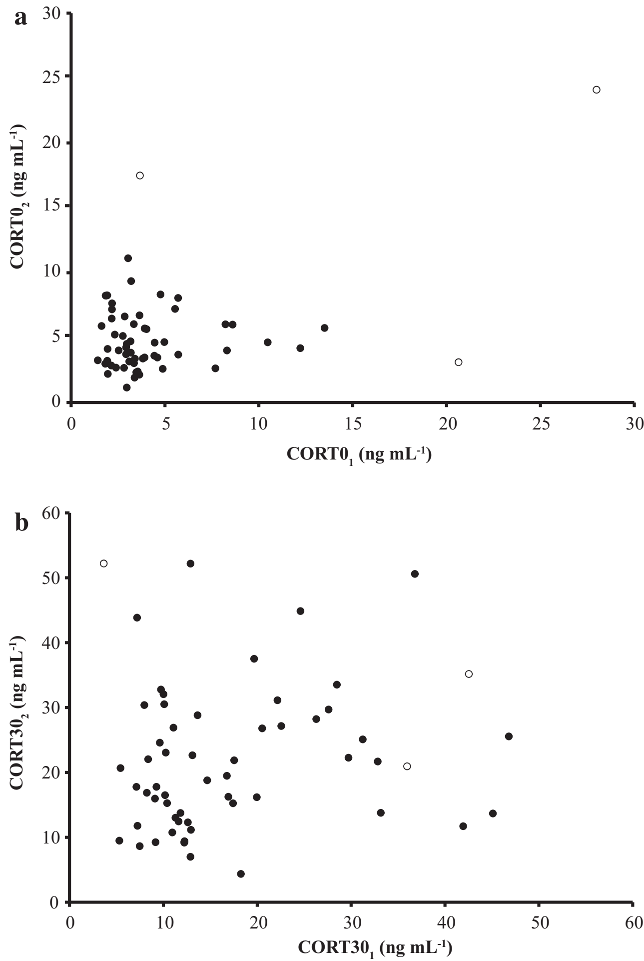


Fig. 2. Relationship between CORT0 at first versus second capture (a). Relationship between CORT30 at first versus second capture (b). The three individual birds that had CORT0 values that were statistical outliers are indicated in open circles in both graphs.

concentrations of stress-induced CORT (Table 2). The decomposition of this phenotypic correlation revealed that this linkage between the two traits was not driven by an among-individual correlation (Fig. 3d; Table 2)—a result that is congruent with the absence of significant among-individual variance in both CORT0 and CORT30. Instead, this phenotypic correlation was driven by a within-individual correlation, which can be illustrated as deviations from individual mean values for CORT0 versus CORT30 (Fig. 4). This means that an individual's change in CORT0 across captures is positively correlated with its change in CORT30 across that same sampling period.

4. Discussion

4.1. Repeatability and levels of variance

The amount of variance within individuals greatly exceeded that observed among individuals for both initial and stress-induced plasma CORT—thus CORT0 and CORT30 did not exhibit significant repeatability. If repeatabilities in these traits are as low as we found in our sample ($r < 0.3$), we estimate that our sample size (58 birds sampled 2–4 times each) provided moderate

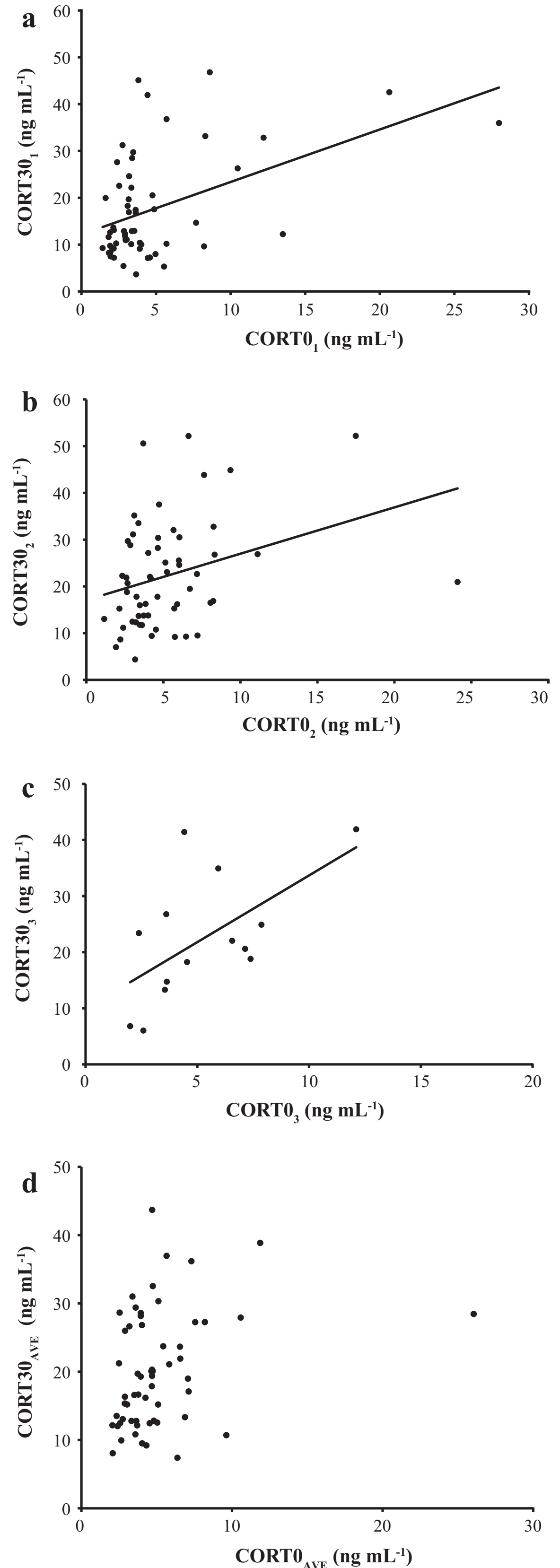


Fig. 3. Correlations between CORT0 and CORT30 at first (a), second (b) and third (c) captures, and the correlation between average (per individual) values (d). Lines represent a Pearson's correlation of the observed data.

Table 2
Bivariate mixed-effects models of the correlations of CORT0 versus CORT30 at the phenotypic, among- and within-individual levels. Statistical outliers ($n = 3$) were included in all models and were either fitted as a fixed effect (*) or not.

Model	Phenotypic correlation				Among individual correlation				Within individual correlation			
	Correlation	SE	Chi^2	P	Correlation	SE	Chi^2	P	Correlation	SE	Chi^2	P
1	0.477	0.050	71.19	3×10^{-17}	0.183	0.386	0.116	0.73	0.564	0.078	28.24	1×10^{-7}
1*	0.455	0.051	62.25	3×10^{-15}	0.122	0.546	0.020	0.89	0.527	0.084	20.31	7×10^{-6}

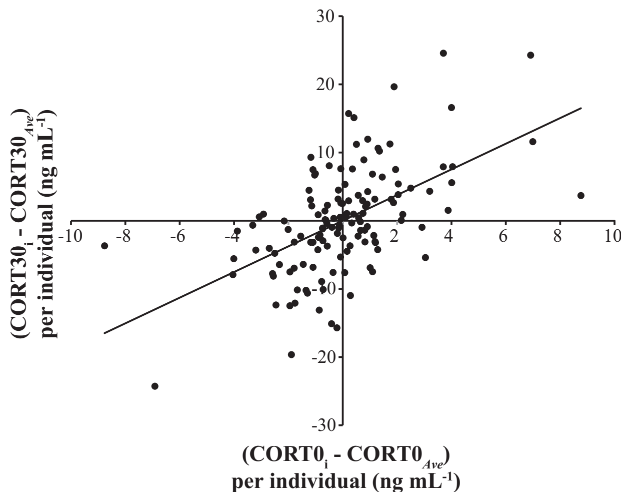


Fig. 4. Deviation from average per individual in the correlation between CORT0 and CORT30 ($n = 58$ birds; 132 captures).

statistical power (power ≈ 0.6) to detect significant departures from zero. On the other hand, if repeatabilities are high ($r > 0.5$), we estimate that our sample size would have provided ample power to detect this (power ≈ 0.9) (see power analyses in Dingemanse and Dochtermann, 2013). Our results therefore indicate that circulating concentrations of CORT0 and CORT30 in wild great tits contain relatively little among-individual variation at best. Given our study's large sample size relative to other published work, and the small degree of measurement error, this means that future studies will likely require considerable sample sizes to detect repeatability, if present, in these glucocorticoid measures under wild conditions (Dingemanse and Dochtermann, 2013; Martin et al., 2010; Wolak et al., 2011). In contrast, we had ample statistical power for estimating the within-individual variance in CORT0 and CORT30 and the within-individual correlation between these two traits.

The lack of repeatability in both CORT0 and CORT30 also implies the absence of individual differences in plasticity; in other words, since both measures failed to exhibit repeatability, it is not possible for there to be repeatable individual differences in the change in CORT concentrations (CORT30 minus CORT0; "CORTchange"), which is often calculated and examined as a separate trait. We advise against this practice for two reasons: (1) repeatability in either one or both of the measured variables (CORT0 and CORT30) could generate apparent repeatability in the derived variable (CORTchange); and (2) performing 'statistics on statistics' can inflate type I errors because the uncertainty around the estimate of an individual's plasticity is discarded. For these reasons, and others (see Romero, 2004), we advocate that future studies partition the variance and correlations in each of the measured variables separately, as we have done here.

Our study depended on recaptures of wild birds, and our temporal re-sampling scheme was thus necessarily distributed across a span of time (range: 3–82 days). These methods might help explain why our results exhibit lower repeatabilities compared to captive studies in the same species that had shorter intervals across observations and greater control over external variables (Cockrem and Silverin, 2002b). Because this study was designed to estimate repeatabilities under wild conditions, we estimated 'agreement repeatabilities'—we did not adjust our estimates for the effects of uncontrolled environmental variables (i.e., 'adjusted repeatabilities'; Nakagawa and Schielzeth, 2010; Westneat et al., 2011). While there are instances in which adjusted repeatabilities can be informative, the present study was designed to examine trait variance and correlations under conditions in which selection is actually acting, and selection does not act on 'adjusted phenotypes.'

Conducting these types of studies in free-living animals, however, is critical because the stress of captivity can influence glucocorticoid profiles (Calisi and Bentley, 2009) and other labile traits such as behavior (Niemelä and Dingemanse, 2014), even after protracted periods of acclimatization to captivity (Marra et al., 1995). This indicates that an organism's physical environment can influence its endocrine state, and therefore the repeatability of that state. Nevertheless, our results are generally consistent with previous work, including from captive bird studies. Romero and Reed (2008) found that within-individual variation was generally higher than among-individual variation ($r < 0.5$) in initial CORT in captive house sparrows (*Passer domesticus*), and we agree that this finding complicates studies that employ temporally separated sampling schemes, or use single observations of hormone concentrations as a proxy for individuals or populations. We emphasize, however, that comparing across studies using different statistical approaches (Cockrem and Silverin, 2002b; Ouyang et al., 2011; Pottinger et al., 1992; Rensel and Schoech, 2011; Romero and Reed, 2008; Schjolden et al., 2005) is problematic—this is due, principally, to the fact that non-partitioned phenotypic variances and correlations can be driven by variation within-individuals in 'state' (e.g., hunger state or temperature). Whereas the use of mixed models for variance partitioning has become common in fields such as evolutionary quantitative genetics (Roff, 1997; Wilson et al., 2010) and ecology (Bolker et al., 2009), these tools are now rapidly gaining popularity behavioral ecology (Dingemanse and Dochtermann, 2013), and this has been instrumental in determining the structure and repeatability of behavioral traits (Dingemanse et al., 2012b; Wolf and Weissing, 2012). Moreover, compared to classical statistical approaches to repeatability (e.g., ANOVA; Lessells and Boag, 1987), mixed models have some important advantages: (1) they allow for the direct estimation of within- and among-individual variances thereby identifying whether differences in repeatabilies among groups are attributable to among-individual variances, within-individual variances or both (Jenkins, 2011; Nakagawa and Schielzeth, 2010); (2) they do not require an equal number of repeated samples per individual (i.e., 'missing cells' are not a problem); and (3) they allow for the calculation of repeatability of traits that have non-Gaussian error distributions (Nakagawa and Schielzeth, 2010). Furthermore, multivariate mixed models permit the portioning of correlations, which provides estimates of the true among-individual correlations that are often of interest. We therefore encourage future studies to employ a repeated

measures sampling scheme and a mixed modeling approach to partition variance and covariance in hormone traits across hierarchical levels (e.g., species, population, individual, observations within the same individual).

We used a standardized stressor procedure common for studies of avian stress. Under more naturalistic circumstances, such as an actual predation attempt, animals might cope behaviorally (Koolhaas et al., 1999) and therefore more quickly turn off their stress response (Rich and Romero, 2005). As a future research direction, we recommend examining the repeatability and correlation of glucocorticoids using shorter duration stressors or those that allow for behavioral coping (Baugh, unpublished). Moreover, it would be informative to measure other components of stress reactivity—a lack of repeatability at initial and 30 min post-stressor time points in great tits does not rule out the possibility that other features of the stress axis exhibit repeatability, such as the timing of stress onset or strength of negative feedback (Baugh, unpublished data), or that inter-specific or seasonal variation underlies disparate findings. Indeed, we have previously shown that these very early and late stages of the stress response are linked with repeatable behavioral traits (Baugh et al., 2013).

4.2. Levels of correlation: are initial and stress-induced CORT separate traits?

We found a strong phenotypic correlation between initial and stress-induced concentrations (Fig. 3a–c; Table 2). Consistent with the lack of among-individual variance, we demonstrated that this phenotypic correlation was driven by a positive within-individual, but not among-individual, correlation. This finding implies that either internal (e.g., nutritional state) or external variables (e.g., social environment) or both varied across observations of the same individual and that these variables modulated the expression of both traits simultaneously. This result provides empirical evidence that initial and stress-induced concentrations of CORT are strongly inter-dependent. This inter-dependence is presumably due to shared mechanisms responsible for co-regulating baseline and stress-induced concentrations (e.g., co-activation of MR and GR receptors) and a sensitivity of these mechanisms to the same environmental or internal factors. Temperature and photoperiod, for instance, might simultaneously influence glucocorticoid concentrations at baseline and during the stress response (Romero et al., 2000; Breuner et al., 1999), although such co-modulatory effects at both the within- and among-individual levels have not, to our knowledge, been demonstrated. Moreover, photoperiod and time of day were largely controlled in the present study, and the influence of temperature that has been described previously is relatively weak (Romero et al., 2000). Therefore, in isolation, these two environmental variables are unlikely to explain the strong positive within-individual correlation observed here. An alternative possibility that deserves future study is that CORT0 estimates in fact fail to capture true baseline concentrations, due for example, to stressful events occurring prior to capture. If CORT0 concentrations are 'contaminated' by the stress response, we might expect a positive within-individual correlation between CORT0 and CORT30 because the CORT0 concentration would partly measure what would normally be in the CORT30 component. Given the low concentrations of CORT that we observed at the initial time point (ca. 5 ng mL^{-1}), it is unlikely that this potential confound explains the strong within-individual correlations observed in the present study. Therefore, in summary, we think an unmeasured environmental variable or set of variables underpins the expression of both CORT0 and CORT30 due to shared mechanisms underlying these two traits.

If CORT0 and CORT30 were both expressions of a repeatable yet labile trait, we have specific predictions about their correlation structure: they should be correlated in the same way (i.e., sign) at every level where significant variance is observed (i.e., both among- and within-individuals for a repeatable and labile trait; reviewed in Araya-Ajoy and Dingemanse, 2013). In the present study, this prediction is confirmed: since there is no variance among individuals, the two traits also cannot be correlated at the among-individual level (cf. covariance requires variance), so this level can be ignored. However, a strong correlation exists at the only level that contains variation (cf. within-individual) implying that they are expressions of the same trait. Note, however, that if observations of the two traits are misassigned, which would occur if 'baseline' samples include stress contamination, the validity of this interpretation is violated.

5. Conclusions

In the present study we demonstrated two important aspects of the glucocorticoid system. First, concentrations of glucocorticoid hormones are highly variable (i.e., not repeatable) within wild great tits. This result advises against interpreting a dataset of single observations of hormone concentrations across individuals as 'individual differences' (Guimont and Wynne-Edwards, 2006; Love et al., 2004; Williams, 2008). Second, two measures of the glucocorticoid system (initial and stress-induced CORT) are strongly and positively correlated within the individual. This means that the concentration of CORT0 at a given time point has some predictive value for the immediately subsequent stress induced concentration, but little or no predictive value for later (or earlier) instances of either measure in the individual's lifetime. This second conclusion further suggests against treating initial and stress-induced concentrations as independent traits. These results emphasize the need to better understand the forces acting within the individual that generate such considerable variation, and cautions against drawing inferences about among-individual differences in the absence of a repeated measures sampling design and an appropriate partitioning of the variance components. The same caution therefore applies to inferring relationships between such 'snapshots' of hormone concentrations and longer-term life history traits, such as survival—a problem under scrutiny in animal (Dingemanse et al., 2010) and human studies of stress physiology (Hruschka et al., 2005). Given the disparate findings for hormone-fitness relationships observed in free-living animal populations (Blas et al., 2007; Bonier et al., 2009; Cabezas et al., 2007; MacDougall-Shackleton et al., 2009; McGlothlin et al., 2010; Patterson et al., 2014; Pride, 2005), it will be important moving forward to identify experimental methods that permit control over possible correlated environmental or developmental influences (see Dingemanse et al., 2010; MacDougall-Shackleton et al., 2013).

That said, a focus beyond circulating hormones that includes other components of endocrine systems (e.g., binding globulins, receptor densities) is also critically needed (Williams, 2008). For example, it is conceivable that despite highly variable circulating concentrations of hormones, covariation in receptor densities results in system-wide repeatability at the target tissue. Conversely, individuals may exhibit similarity in circulating concentrations, but due to differences in end organ sensitivity, exhibit differences in responses. Future endeavors along these lines would contribute significantly to our understanding of the evolution of endocrine systems.

We believe our results are consistent with the understanding of endocrine systems as dynamic mediators that help individual organisms maintain stasis in a background of environmental change (McEwen and Wingfield, 2003), and are incongruent with the concept of hormone concentrations as highly repeatable traits. Indeed, recent work in mammals has demonstrated that unraveling HPA-behavior relationships may require methods that permit

measurement of the fine-scale temporal dynamics of the stress axis (Sarabdjitsingh et al., 2010). We encourage future studies to consider hormone concentrations as labile traits that mediate those phenotypic characters available for selection to act on (e.g., behavior, performance, ornamentation). Modeling the relationships across these levels of phenotypic organization using a reaction norm approach will help us better understand how such regulatory systems contribute to functional consequences at the individual level (Arnold, 1983; McGlothlin and Ketterson, 2008).

Acknowledgments

We thank S. Davidson for assisting with fieldwork. This manuscript was improved following feedback from M. Firke, R. Senft, A. Lauder and anonymous reviewers. This work was supported by the Alexander von Humboldt Foundation (ATB), the Max Planck Society (NJD, MH) and NIOO-KNAW and was carried out in accordance with the animal ethical committee of the Royal Dutch Academy of Sciences (DEC-KNAW) under protocol NIOO 10.05 to KVO.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ygcen.2014.08.014.

References

Almasi, B., Roulin, A., Jenni, L., 2013. Corticosterone shifts reproductive behaviour towards self-maintenance in the barn owl and is linked to melanin-based coloration in females. Horm. Behav. 64, 161–171.
Araya-Ajoy, Y.G., Dingemanse, N.J., 2013. Characterizing behavioural 'characters': an evolutionary framework. Proc. R. Soc. B 281, 20132645.
Arnold, S.J., 1983. Morphology, performance and fitness. Am. Zool. 23, 347–361.
Baugh, A.T., Schaper, S.V., Hau, M., Cockrem, J.F., de Goede, P., van Oers, K., 2012. Corticosterone responses differ between lines of great tits (*Parus major*) selected for divergent personalities. Gen. Comp. Endocrinol. 175, 488–494.
Baugh, A.T., van Oers, K., Naguib, M., Hau, M., 2013. Initial reactivity and magnitude of the acute stress response associated with personality in wild great tits (*Parus major*). Gen. Comp. Endocrinol. 189, 96–104.
Blas, J., Bortolotti, G.R., Tella, J.L., Baos, R., Marchant, T.A., 2007. Stress response during development predicts fitness in a wild, long lived vertebrate. Proc. Natl. Acad. Sci. U.S.A. 104, 8880–8884.
Bókony, V., Lendvai, Á.Z., Liker, A., Angelier, F., Wingfield, J.C., Chastel, O., 2009. Stress response and the value of reproduction: are birds prudent parents? Am. Nat. 173, 589–598.
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 Ecol. Evol. 24, 127–135.
Bonier, F., Moore, I.T., Martin, P.R., Robertson, R.J., 2009. The relationship between fitness and baseline glucocorticoids in a passerine bird. Gen. Comp. Endocrinol. 163, 208–213.
Breuner, C.W., Hahn, T.P., 2003. Integrating stress physiology, environmental change, and behavior in free-living sparrows. Horm. Behav. 43, 115–123.
Breuner, C.W., Wingfield, J.C., Romero, L.M., 1999. Diel rhythms of basal and stress-induced corticosterone in a wild, seasonal vertebrate, Gambel's white-crowned sparrow. J. Exp. Zool. 284, 334–342.
Butler, L.K., Bisson, I.A., Hayden, T.J., Wikelski, M., Romero, L.M., 2009. Adrenocortical responses to offspring-directed threats in two open-nesting birds. Gen. Comp. Endocrinol. 162, 313–318.
Cabezas, S., Blas, J., Marchant, T.A., Moreno, S., 2007. Physiological stress levels predict survival probabilities in wild rabbits. Horm. Behav. 51, 313–320.
Calisi, R.M., Bentley, G.E., 2009. Lab and field experiments: are they the same animal? Horm. Behav. 56, 1–10.
Carere, C., Groothuis, T.G.G., Mostl, E., Daan, S., Koolhaas, J.M., 2003. Fecal corticosteroids in a territorial bird selected for different personalities: daily rhythm and the response to social stress. Horm. Behav. 43, 540–548.
Carere, C., Caramaschi, D., Fawcett, T.W., 2010. Covariation between personalities and individual differences in coping with stress: converging evidence and hypotheses. Curr. Zool. 56, 728–740.
Cockrem, J.F., Silverin, B., 2002a. Sight of a predator can stimulate a corticosterone response in the great tit (*Parus major*). Gen. Comp. Endocrinol. 125, 248–255.
Cockrem, J.F., Silverin, B., 2002b. Variation within and between birds in corticosterone responses of great tits (*Parus major*). Gen. Comp. Endocrinol. 125, 197–206.
Dingemanse, N.J., Dochtermann, N.A., 2013. Quantifying individual variation in behaviour: mixed-effect modelling approaches. J. Anim. Ecol. 82, 39–54.
Dingemanse, N.J., Dochtermann, N.A., 2014. Individual behaviour: behavioural ecology meets quantitative genetics. In: Charmantier, A., Garant, D., Kruuk, L.E.B. (Eds.), Quantitative Genetics in the Wild. Oxford University Press, in press, http://dx.doi.org/10.1093/acprof:oso/9780199674237.003.0004.
Dingemanse, N.J., Edelaar, P., Kempenaers, B., 2010. Why is there variation in baseline glucocorticoid levels? Trends Ecol. Evol 25, 261–262.
Dingemanse, N.J., Dochtermann, N.A., Nakagawa, S., 2012a. Defining behavioural syndromes and the role of "syndrome deviation" to study its evolution. Behav. Ecol. Sociobiol. 66, 1543–1548.
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*. J. Anim. Ecol. 81, 116–126.
Dochtermann, N.A., 2011. Testing Cheverud's conjecture: behavioral correlations and behavioral syndromes. Evolution 65, 1814–1820.
Dohm, M.R., 2002. Repeatability estimates do not always set an upper limit to heritability. Funct. Ecol. 16, 273–280.
Duckworth, R.A., Sockman, K.W., 2012. Proximate mechanisms of behavioural inflexibility: implications for the evolution of personality traits. Funct. Ecol. 26, 559–566.
Eikenaar, C., Husak, J., Escallón, C., Moore, I.T., 2012. Variation in testosterone and corticosterone in amphibians and reptiles: relationships with latitude, elevation, and breeding season length. Am. Nat. 180, 642–654.
Eilam, D., Dayan, T., Ben-Eliyahu, S., Schulman, I., Shefer, G., Hendrie, C.A., 1999. Differential behavioural and hormonal responses of voles and spiny mice to owl calls. Anim. Behav. 58, 1085–1093.
Goymann, W., Moore, I.T., Scheuerlein, A., Hirschenhauser, K., Grafen, A., Wingfield, J.C., 2004. Testosterone in tropical birds: effects of environmental and social factors. Am. Nat. 164, 327–334.
Guimont, F.S., Wynne-Edwards, K.E., 2006. Individual variation in cortisol responses to acute 'on-back' restraint in an outbred hamster. Horm. Behav. 50, 252–260.
Harri, M., Mononen, L., Ahola, I., Plyusnina, Rekilä T., 2003. Behavioural and physiological differences between silver foxes selected and not selected for domestic behaviour. Anim Welf. 12, 305–314.
Hau, M., 2007. Regulation of male traits by testosterone: implications for the evolution of vertebrate life histories. BioEssays 29, 133–144.
Hau, M., Ricklefs, R.E., Wikelski, M., Lee, K.A., Brawn, J.D., 2010. Corticosterone, testosterone and life-history strategies of birds. Proc. R. Soc. B 277, 3203–3212.
Heideman, P.D., 2004. Top-down approaches to the study of natural variation in complex physiological pathways using the white-footed mouse (*Peromyscus leucopus*) as a model. ILAR J. 45, 4–13.
Hruschka, D.J., Kohrt, B.A., Worthman, C.M., 2005. Estimating between- and within-individual variation in cortisol levels using multilevel models. Psychoneuroendocrinol. 30, 698–714.
Jawor, J.M., McGlothlin, J.W., Casto, J.M., Greives, T.J., Snajdr, E.A., Bentley, G.E., Ketterson, E.D., 2006. Seasonal and individual variation in response to GnRH challenge in male dark-eyed juncos (*Junco hyemalis*). Gen. Comp. Endocrinol. 149, 182–189.
Jenkins, S.H., 2011. Sex differences in repeatability of food-hoarding behaviour of kangaroo rats. Anim. Behav. 81, 1155–1162.
Ketterson, E.D., Nolan Jr., V., Cawthorn, M.J., Parker, P.G., Ziegenfus, C., 1996. Phenotypic engineering: using hormones to explore the mechanistic and functional bases of phenotypic variation in nature. Ibis 138, 70–86.
Ketterson, E.D., Atwell, J.W., McGlothlin, J.W., 2009. Phenotypic integration and independence: hormones, performance, and response to environmental change. Integr. Comp. Biol. 49, 365–379.
Koolhaas, J.M., Korte, S.M., De Boer, S.F., Van Der Vegt, B.J., Van Reenen, C.G., Hopster, H., De Jong, I.C., Ruis, M.A.W., Blokhuis, H.J., 1999. Coping styles in animals: current status in behavior and stress-physiology. Neurosci. Biobehav. Rev. 23, 925–935.
Kralj-Fiser, S., Scheiber, I.B.R., Blejec, A., Moestl, E., Kotrschal, K., 2007. Individualities in a flock of free-roaming greylag geese: behavioral and physiological consistency over time and across situations. Horm. Behav. 51, 239–248.
Lande, R., Arnold, S.J., 1983. The measurement of selection on correlated characters. Evolution 37, 1210–1226.
Landys, M.M., Ramenofsky, M., Wingfield, J.C., 2006. Actions of glucocorticoids at a seasonal baseline as compared to stress-related levels in the regulation of periodic life processes. Gen. Comp. Endocrinol. 148, 132–149.
Landys, M.M., Goymann, W., Schwabl, I., Trapschuh, M., Slagsvold, T., 2010. Impact of season and social challenge on testosterone and corticosterone levels in a year-round territorial bird. Horm. Behav. 58, 317–325.
Lessells, C.M., Boag, P.T., 1987. Auk 104, 116–121.
Love, O.P., Breuner, C.W., Vézina, F., Williams, T.D., 2004. Mediation of a corticosterone-induced reproductive conflict. Horm. Behav. 46, 59–65.
Lynn, S.E., Prince, L.E., Phillips, M.M., 2010. A single exposure to an acute stressor has lasting consequences for the hypothalamo-pituitary-adrenal response to stress in free-living birds. Gen. Comp. Endocrinol. 165, 337–344.
MacDougall-Shackleton, S.A., Dindia, L., Newman, A.E.M., Potvin, D.A., Stewart, K.A., MacDougall-Shackleton, E.A., 2009. Stress, song and survival in sparrows. Biol. Lett. 5, 746–748.
MacDougall-Shackleton, S.A., Schmidt, K.L., Furlonger, A.A., MacDougall-Shackleton, E.A., 2013. HPA axis regulation, survival, and reproduction in free-living sparrows: functional relationships or developmental correlations? Gen. Comp. Endocrinol. 190, 188–193.

Marra, P.P., Lampe, K.T., Tedford, B.L., 1995. Plasma corticosterone levels in two species of *Zonotrichia* sparrows under captive and free-living conditions. Wilson Bull. 107, 296–305.
Martin, J.G.A., Nussey, D.H., Wilson, A.J., Réale, D., 2010. Measuring individual differences in reaction norms in field and experimental studies: a power analysis of random regression models. Methods Ecol. Evol. 2, 362–374.
McEwen, B.S., Wingfield, J.C., 2003. The concept of allostasis in biology and biomedicine. Horm. Behav. 43, 2–15.
McGlothlin, J.W., Ketterson, E.D., 2008. Hormone-mediated suites as adaptations and evolutionary constraints. Philos. Trans. R. Soc. Lond. B 363, 1611–1620.
McGlothlin, J.W., Whittaker, D.J., Schrock, S.E., Gerlach, N.M., Jawor, J.M., Snajdr, E.A., Ketterson, E.D., 2010. Natural selection on testosterone production in a wild songbird population. Am. Nat. 175, 687–701.
Meyer, K., 1992. Variance components due to direct and maternal effects for growth traits of australian beef cattle. Livestock Prod. Sci. 31, 179–204.
Nakagawa, S., Schielzeth, H., 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biol. Rev. 85, 935–956.
Niemelä, P.T., Dingemanse, N.J., 2014. Artificial environments and the study of 'adaptive' personalities. Trends Ecol. Evol. 29, 245–247.
Ouyang, J.Q., Hau, M., Bonier, F., 2011. Within seasons and among years: when are corticosterone levels repeatable? Horm. Behav. 60, 559–564.
Patterson, S.H., Hahn, T.P., Cornelius, J.M., Breuner, C.W., 2014. Natural selection and glucocorticoid physiology. J. Evol. Biol. 27, 259–274.
Pottinger, T.G., Pickering, A.D., Hurley, M.A., 1992. Consistency in the stress response of individuals of two strains of rainbow trout, *Oncorhynchus mykiss*. Aquaculture 103, 275–289.
Pride, R.E., 2005. High faecal glucocorticoid levels predict mortality in ring-tailed lemurs (*Lemur catta*). Biol. Lett. 1, 60–63.
Rensel, M.A., Schoech, S.J., 2011. Repeatability of baseline and stress-induced corticosterone levels across early life stages in the Florida scrub-jay (*Aphelocoma coerulescens*). Horm. Behav. 59, 497–502.
Rich, E.L., Romero, L.M., 2005. Exposure to chronic stress downregulates corticosterone responses to acute stressors. Am. J. Physiol. 288, R1628–R1636.
Roff, D.A., 1997. Evolutionary Quantitative Genetics. Chapman and Hall, New York.
Romero, L.M., 2004. Physiological stress in ecology: lessons from biomedical research. Trends Ecol. Evol. 19, 249–255.
Romero, L.M., Reed, J.M., 2005. Collecting baseline corticosterone samples in the field: is under 3 min good enough? Comp. Biochem. Physiol. 140, 73–79.
Romero, L.M., Reed, J.M., 2008. Repeatability of baseline corticosterone concentrations. Gen. Comp. Endocrinol. 156, 27–33.
Romero, L.M., Romero, R.C., 2002. Corticosterone responses in wild birds: the importance of rapid initial sampling. Condor 104, 129–135.
Romero, L.M., Ramenofsky, M., Wingfield, J.C., 1997. Season and migration alter corticosterone response to capture and handling in an Arctic migrant, the White-crowned sparrow (*Zonotrichia leucophrys gambelii*). Comp. Biochem. Physiol. 116C, 171–177.
Romero, L.M., Reed, J., Wingfield, J.C., 2000. Effects of weather on corticosterone responses in wild free-living passerine birds. Gen. Comp. Endocrinol. 118, 113–122.
Sarabdjitsingh, R.A., Conway-Campbell, B.L., Leggett, J.D., Waite, E.J., Meijer, O.C., de Kloet, E.R., Lightman, S.L., 2010. Stress responsiveness varies over the ultradian glucocorticoid cycle in a brain-region-specific manner. Endocrinology 151, 5369–5379.
Schjolden, J., Stoskhus, A., Winberg, S., 2005. Does individual variation in stress responses and agonistic behavior reflect divergent stress coping strategies in juvenile rainbow trout? Physiol. Biochem. Zool. 78, 715–723.
Schmidt, K.L., Furlonger, A.A., Lapierre, J.M., MacDougall-Shackleton, E.A., MacDougall-Shackleton, S.A., 2012. Regulation of the HPA axis is related to song complexity and measures of phenotypic quality in song sparrows. Horm. Behav. 61, 652–659.
Van Duyse, E., Pinxten, R., Darras, V.M., Arckens, L., Eens, M., 2004. Opposite changes in plasma testosterone and corticosterone levels following a simulated territorial challenge in male great tits. Behaviour 141, 451–467.
Van Oers, K., Buchanan, K.L., Thomas, T.E., Drent, P.J., 2011. Correlated response to selection of testosterone levels and immunocompetence in lines selected for avian personality. Anim. Behav. 81, 1055–1061.
Via, S., Gomulkiewicz, R., Dejong, G., Scheiner, S.M., Schlichting, C.D., Vantienderen, P.H., 1995. Adaptive phenotypic plasticity – consensus and controversy. Trends Ecol. Evol. 10, 212–217.
Vitousek, M.N., Rubenstein, D.R., Nelson, K.N., Wikelski, M., 2008. Are hotshots always hot? A longitudinal study of hormones, behavior, and reproductive success in male marine iguanas. Gen. Comp. Endocrinol. 157, 227–232.
Wada, H., Salvante, K.G., Stables, C., Wagner, E., Williams, T.D., Breuner, C.W., 2008. Adrenocortical responses in zebra finches (*Taeniopygia guttata*): individual variation, repeatability, and relationship to phenotypic quality. Horm. Behav. 53, 472–480.
Westneat, D.F., Hatch, M.I., Wetzel, D.P., Ensminger, A.L., 2011. Indiviudual variation in parental care reaction norms: integration of personality and plasticity. Am. Nat. 178, 652–667.
Westneat, D.F., Wright, J., Dingemanse, N.J., 2014. The biology hidden inside residual within-individual phenotypic variation. Biological Reviews, in press, http://dx.doi.org/10.1111/brv.12131.
While, G.M., Isaksson, C., McEvoy, J., Sinn, D.L., Komdeur, J., Wapstra, E., Groothuis, T.G.G., 2010. Repeatable intra-individual variation in plasma testosterone concentration and its sex-specific link to aggression in a social lizard. Horm. Behav. 58, 208–213.
Williams, T.D., 2008. Individual variation in endocrine systems: moving beyond the 'tyranny of the Golden Mean'. Philos. Trans. R. Soc. B. 363, 1687–1698.
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. J. Anim. Ecol. 79, 13–26.
Wingfield, J.C., Lewis, D.M., 1993. Hormonal and behavioural responses to simulated territorial intrusions in the cooperatively breeding white-browed sparrow weaver, *Plocepasser mahali*. Anim. Behav. 45, 1–11.
Wingfield, J.C., Smith, J.P., Farner, D.S., 1982. Endocrine responses of white-crowned sparrows to environmental stress. Condor 84, 399–409.
Wingfield, J.C., Maney, D.L., Breuner, C.W., Jacobs, J.D., Lynn, S., Ramenofsky, M., Richardson, R.D., 1998. Ecological bases of hormone-behavior interactions: the "Emergency Life History Stage". Am. Zool. 38, 191–206.
Wolak, M.E., Fairbairn, D.J., Paulsen, Y.R., 2011. Guidelines for estimating repeatability. Methods Ecol. Evol. 3, 129–137.
Wolf, M., Weissing, F.J., 2012. Animal personalities: consequences for ecology and evolution. Trends Ecol. Evol. 27, 452–461.
Zera, A.J., Harshman, L.G., Williams, T.D., 2007. Evolutionary endocrinology: the developing synthesis between endocrinology and evolutionary genetics. Annu. Rev. Ecol. Evol. Syst. 38, 793–817.