Molecular and Cellular Endocrinology 500 (2020) 110640

Contents lists available at ScienceDirect

Molecular and Cellular Endocrinology

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

Do hormone manipulations reduce fitness? A meta-analytic test of the Optimal Endocrine Phenotype Hypothesis

Frances Bonier[a,*], Robert M. Cox[b]

[a] Department of Biology, Queen's University, Kingston, Ontario, K7L 3N6, Canada
[b] Department of Biology, University of Virginia, Charlottesville, VA, 22904, USA

ARTICLE INFO

Keywords:
Adaptive plasticity
Androgens
Glucocorticoids
Natural selection
Optimality
Meta-analysis

ABSTRACT

Endocrine traits (e.g., circulating hormone concentrations, receptor expression) can vary considerably among individuals within populations. Here, we develop two evolutionary hypotheses to explain this variation. Under the Optimal Endocrine Phenotype Hypothesis, adaptive plastic responses to environmental variation generate individual variation in endocrine traits and allow individuals to express near-optimal endocrine phenotypes. In contrast, under the Ongoing Selection Hypothesis, individual variation in endocrine traits reflects varying adaptive value, with some individuals expressing suboptimal phenotypes that are selected against. These two hypotheses generate distinct predictions for the effects of hormone manipulations on fitness. Under the Optimal Endocrine Phenotype Hypothesis, all hormone manipulations should incur fitness costs, whereas under the Ongoing Selection Hypothesis, manipulating endocrine phenotypes toward a putative optimum should increase fitness. Using a meta-analysis of findings from experimental field studies that involved manipulation of circulating glucocorticoids or androgens and measurement of fitness effects, we test and find some support for the Optimal Endocrine Phenotype Hypothesis. On average, fitness was reduced across 97 estimates of the effects of experimental hormone manipulations on fitness. However, the fitness effects of glucocorticoid manipulations varied with the sex of the individuals being studied. Fitness was more uniformly reduced by glucocorticoid manipulations in males and when both sexes were considered together. In females, effects on fitness varied from highly positive to highly negative. The effects of androgen manipulations varied across males and females, and depending upon whether fitness was estimated using measures of reproductive success or survival. Reproductive success was consistently decreased by androgen manipulation in females, but was increased almost as often as it was decreased across experiments in males. When survival was estimated as a component of fitness, it was fairly uniformly compromised by exogenous androgens in males. This variation in fitness effects of hormone manipulations across sexes and fitness metrics is consistent with the expectation that hormones differentially regulate life-history investment and that optimal endocrine phenotypes differ between males and females. Overall, our meta-analysis provides some support for the Optimal Endocrine Phenotype Hypothesis, but we await direct tests of the Ongoing Selection Hypothesis to determine the degree to which individual variation in endocrine traits continues to be shaped by natural selection.

1. Introduction

Recent decades have seen a dramatic increase in interest in explaining among-individual variation in endocrine traits (e.g., circulating hormone concentrations, receptor expression, endocrine axis regulation) (Bonier et al., 2009b; Hau et al., 2016; McGlothlin et al., 2007, 2010; Patterson et al., 2014; Taff and Vitousek, 2016; Tschirren et al., 2014). Fundamental to this interest is the goal of understanding the potential adaptive value of this variation (Ketterson et al., 1996; Taff and Vitousek, 2016; Williams, 2007). Many of us take an adaptationist approach to interpreting highly conserved endocrine responses and the regulation of endocrine axes (Love and Williams, 2008; Råberg et al., 1998; Williams, 2007). For example, given that the vast majority of individuals across vertebrate species increase circulating concentrations of glucocorticoid hormones when confronted with challenges; the consistency of this pattern across disparate taxa suggests that the response is adaptive. Increasing circulating glucocorticoid concentrations in response to challenges would be expected to increase an individual's fitness, relative to the alternative of not having an intact or appropriate glucocorticoid response. If these plastic responses to

* Corresponding author.
E-mail address: bonierf@queensu.ca (F. Bonier).

https://doi.org/10.1016/j.mce.2019.110640
Received 2 July 2019; Received in revised form 3 November 2019; Accepted 5 November 2019
Available online 09 November 2019
0303-7207/

environmental variation are adaptive, we might expect that differences in baseline glucocorticoid concentrations among individuals within a population primarily reflect variation in the degree of challenge that each individual is facing (or faced in the past). We could apply a similar, adaptationist logic to an array of endocrine traits, making sense of the variation we see in these traits among individuals.

A common approach to understanding individual variation in endocrine traits is to quantify trait-fitness correlations and infer how selection is acting on that variation (McGlothlin et al., 2010; Patterson et al., 2014; Tschirren et al., 2014). This approach implicitly assumes that some individuals express suboptimal or maladapted trait values relative to others, and that this variation in the adaptive value of individual phenotypes might be revealed by describing the covariance between endocrine trait values and fitness. Individuals with higher fitness might have endocrine trait values that are closer to optimal and favored by natural selection (Lande and Arnold, 1983; Schluter, 1988). Under this scenario, rather than applying an adaptationist assumption to explain individual variation, we are instead focusing on global optima for the endocrine trait of interest, and inferring that the endocrine phenotypes of some individuals are better matched to those optima than those of other individuals in the same population. These two perspectives differ in their fundamental explanations of individual variation in endocrine phenotypes – one ascribes this variation primarily to adaptive plastic responses to environmental variation (hereafter the *Optimal Endocrine Phenotype Hypothesis*), while the other ascribes individual variation to differences in the adaptive value of endocrine phenotypes (hereafter the *Ongoing Selection Hypothesis*). While we present these hypotheses as alternatives, we should note they are opposite ends of a continuum, with individual variation reflecting optimal trait expression at one end and maladaptation (or constraints limiting the optimality of trait expression) at the other end of the spectrum (Ketterson and Nolan, 1999). Natural variation in biological populations will likely fall somewhere between these extremes, but the extent to which optimality or maladaptation and constraints explain individual variation could be revealed through tests of these hypotheses. Although currently available data only support a direct test of the Optimal Endocrine Phenotype Hypothesis, we also discuss and present predictions of the Ongoing Selection Hypothesis to encourage future work that would provide strong tests of both hypotheses.

The basic question, which, when answered, would allow us to distinguish the degree to which these two hypotheses explain individual variation in endocrine phenotypes within a population is: How well do individual endocrine phenotypes match environmentally dictated optima? We can imagine that variation in the environment generates an adaptive endocrine landscape (Arnold et al., 2001; Calsbeek et al., 2012). As with all adaptive landscapes, this optimal endocrine landscape is multidimensional, depending on both abiotic and biotic environmental factors (Calsbeek et al., 2012). For most endocrine traits, this landscape should also shift with an individual's context (e.g., through seasonal, ontogenetic, and life-history transitions), and be dependent on factors such as sex, age, health status, local environment, and past experience (e.g., developmental conditions). Given these many sources of variation shaping the fitness landscape, no two individuals within a population are likely occupying the exact same location on the same multidimensional landscape at the same time (i.e., experiencing the same combination of sex, age, environmental conditions, life-history stage, health status, local environment, and developmental history). If past selection and adaptive plasticity (i.e., adaptive phenotypic flexibility) is sufficient to match the temporally and spatially variable environment, thereby allowing individuals to express near-optimal endocrine phenotypes, then we should expect to see considerable individual variation in endocrine traits. The Optimal Endocrine Phenotype Hypothesis therefore proposes that endocrine variation reflects both past responses to selection as well as continuous, adaptive adjustment of endocrine phenotypes to track shifting environmental optima, which are often variable among individuals. In contrast, the Ongoing Selection Hypothesis would agree that these adaptive landscapes exist, but would assert that most of the variation in individual endocrine phenotypes reflects failure to express optimal trait values.

Importantly, the Optimal Endocrine Phenotype Hypothesis does not preclude the existence of strong correlations between endocrine traits and fitness. Instead, this hypothesis would assert that trait-fitness correlations would be expected, but rather than reflecting ongoing natural selection, they would be driven by adaptive plastic responses to the environment, with different environments being associated with variation in maximal reproduction and survival (Bonier and Martin, 2016; Mitchell-Olds and Shaw, 1987; Price et al., 1988). For example, while increasing concentrations of glucocorticoids might be an adaptive and highly conserved response to resource limitation (Honarmand et al., 2010; Lendvai et al., 2014; Lynn et al., 2003), individuals with fewer resources will still produce fewer offspring than individuals with more resources (Martin, 1987; Price et al., 1988). In this example, elevated glucocorticoids will be associated with lower reproductive success, but not because these glucocorticoid concentrations are selected against. Indeed, they are favored in this challenging environment, but so is restraint in reproductive investment (Bonier and Martin, 2016; Mitchell-Olds and Shaw, 1987). The degree to which endocrine and life history phenotypes shift in response to environmental variation will also depend on life-history strategies, which vary among individuals, sexes, and taxa, and could modulate the fitness effects of hormone manipulations. Importantly, the Optimal Endocrine Phenotype and the Ongoing Selection hypotheses generate distinct, testable predictions. If past selection and adaptive plasticity are the primary drivers of individual variation in an endocrine phenotype (Optimal Endocrine Phenotype Hypothesis), any experimental manipulation of that phenotype ought to decrease fitness (Fig. 1A) (Ketterson et al., 1996). In contrast, if failure to express optimal phenotypes explains individual variation (Ongoing Selection Hypothesis), then shifting an individual's endocrine phenotype toward an optimum ought to increase fitness (Fig. 1B).

Under the Ongoing Selection Hypothesis, we should be able to use a naturally occurring, population-level fitness surface to generate a prediction for the effect of an experimental manipulation on fitness. For example, if higher concentrations of a circulating hormone are correlated with higher reproductive success in females within a population, then experimentally increasing concentrations of that hormone in individuals with low concentrations, within a biologically realistic range that moves them toward (and not beyond) the hypothesized optimal phenotype, should increase reproductive success (Fig. 1B). If selection favors generally increased or decreased hormone concentrations across all individuals (i.e., true directional selection, with the optimal value occurring beyond that of any trait values currently expressed by individuals in the population), then hormone manipulation across all individuals could increase fitness. This latter scenario seems less biologically realistic, however, and so we would need endocrine manipulations only of individuals with sub-optimal trait values (as determined by the hormone-fitness relationship) for a robust test of the Ongoing Selection Hypothesis.

Here, we use a meta-analytic approach to test the Optimal Endocrine Phenotype Hypothesis. As described above, the available data are not suitable for a direct test of the Ongoing Selection Hypothesis, because under that hypothesis we would only expect increased fitness when endocrine traits are deliberately manipulated toward a pre-determined population optimum, which would require detailed information about the shape of the fitness surface in that population. We focus on two endocrine traits: circulating androgen and glucocorticoid concentrations. We selected these two traits because they are frequently manipulated in free-ranging individuals, allowing measurement of fitness consequences in animals exposed to natural agents of selection. We limit our analysis to studies that involved experimental augmentation of circulating hormone concentrations and measurement of a direct fitness component (i.e., reproductive success or

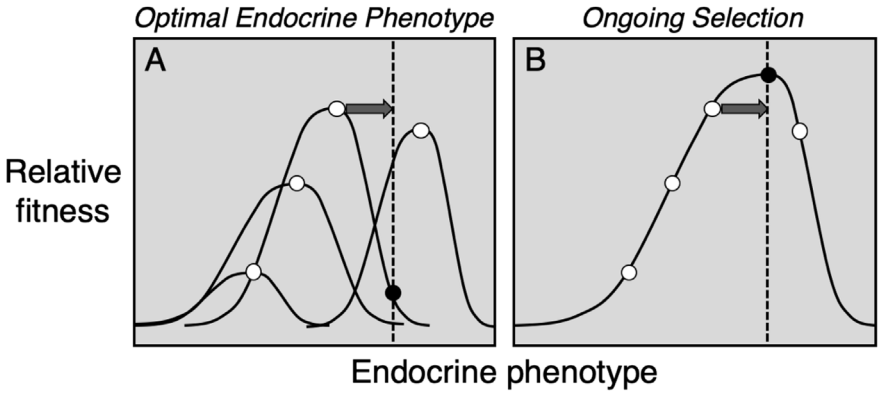

Fig. 1. Illustrations of the Optimal Endocrine Phenotype (A) and Ongoing Selection (B) Hypotheses. Both panels illustrate variation among four individuals (white circles) in an endocrine phenotype (x-axis; hormone concentrations, receptor densities, etc.) and relative fitness (y-axis). Under the Optimal Endocrine Phenotype Hypothesis, we expect that most variation in endocrine phenotypes reflects adaptive responses to environmental variation, and that most individuals express optimal or near-optimal endocrine phenotypes for their environmental context, which varies among individuals and also determines maximal relative fitness. In contrast, under the Ongoing Selection Hypothesis, we expect that much of the variation in endocrine phenotypes reflects variation in the adaptive value of individual traits, with some individuals expressing maladaptive phenotypes. Lines represent fitness surfaces with peaks in panel A denoting multiple, context-dependent optimal endocrine phenotypes, and the single peak in panel B denoting a single, global optimum. The predictions of these two hypotheses for effects of experimental manipulations of an endocrine phenotype differ (arrow and black circles on dashed lines). Under the Optimal Endocrine Phenotype Hypothesis, all manipulations should incur fitness costs, whereas under the Ongoing Selection Hypothesis, manipulation of an individual's phenotype toward the optimum should increase fitness. Under each hypothesis, the magnitude of the fitness effect will be determined by the magnitude of the manipulation (length of arrow), the shape of the underlying fitness landscape (black curves), and the individual's relative position on that landscape.

survival). Following the Optimal Endocrine Phenotype Hypothesis, we predicted that, across studies, fitness should generally be reduced by hormone manipulations.

2. Methods

We reviewed the literature to identify empirical studies that met three criteria. First, they needed to involve experimental manipulation of circulating androgen or glucocorticoid concentrations. These manipulations most often involve subcutaneous implants (silastic, pellet, or osmotic pump), but we also included intraperitoneal, oral, and transdermal dosing. We did not include manipulations of embryonic endocrine phenotypes (e.g., *in ovo* manipulations). Second, we included studies that reported a measure of fitness for both treatment and control groups, and restricted our analyses to studies that assessed reproductive success or survival, excluding those that only reported less direct fitness proxies (e.g., growth or body mass). Third, we only included studies that were conducted with free-ranging animals, with fitness estimated in the field, where study subjects were exposed to natural agents of selection.

Our aim was to assemble a robust dataset for use in testing the Optimal Endocrine Phenotype Hypothesis. Although we conducted a thorough review of the literature, our dataset might not be comprehensive. We conducted database searches (Web of Science, Google Scholar) to identify candidate studies, using various combinations of the following search terms: glucocorticoid, cortisol, corticosterone, cort*, testosterone, androgen, experimental manipulation, implant, fitness, reproduction, productivity, survival, performance. We also systematically reviewed all papers citing a review that evaluated correlative relationships between glucocorticoids and fitness (Bonier et al., 2009a), as well as two recent studies that review glucocorticoid manipulation experiments for different purposes (Crossin et al., 2016; Sopinka et al., 2015).

Using this approach, we identified 65 studies that met our criteria: 27 androgen and 38 glucocorticoid manipulation experiments (Supplementary material, Table S1). The Optimal Endocrine Phenotype Hypothesis asserts that any manipulation of the endocrine phenotype will compromise fitness. However, the vast majority of available studies involve experimental augmentation of circulating hormone concentrations. Indeed, we identified only one study meeting our criteria that involved a putative blocking of the action of testosterone (Dey et al., 2010), and so we excluded this result from our analyses, because we would not have sufficient power to determine if hormone blocking had different effects on fitness than augmentation. Results do not differ if the findings from this study are included in the analysis (fitness was marginally, though not significantly, lower in treatment individuals than controls in the original study). We included a glucocorticoid augmentation result from this same study. Some studies reported effects of the hormone manipulation on multiple fitness metrics, or separately for different sexes or age classes, or for multiple treatment groups (e.g., low and high doses), so in total we included 101 estimated fitness effects (39 from androgen manipulations, 62 from glucocorticoid manipulations) from these studies in our initial analyses. We extracted data on focal species, sample size, sexes included, life-history stages included, method of hormone manipulation, type of fitness metrics measured, and the fitness estimates for the treatment and control groups from each study (Supplementary material, Table S1). When fitness metrics were measured but not clearly reported in the manuscripts, we contacted authors to request data. If data were only presented graphically, we used the program WebPlot Digitizer (https://apps.automeris.io/wpd/) to extract data from the figure image. For each study, we compared the mean fitness of individuals in the treatment group to that of individuals in the control group. Some studies separately reported results for both sham-manipulated individuals (e.g., empty implants) as well as unmanipulated control individuals. In these cases, we based our comparisons on sham-manipulated individuals and excluded data from unmanipulated controls. However, several studies only included unmanipulated controls or presented their data pooled across unmanipulated and sham manipulated controls (generally after finding no differences between those groups), and so, in these cases, we included these fitness estimates from unmanipulated control animals in our analyses.

To estimate the study-specific effect of a hormone manipulation on fitness, we divided the mean fitness of the treatment group by the mean fitness of the control group and then natural-log transformed this ratio, to center it on zero. With this calculation, a negative treatment:control fitness ratio indicates that the hormone manipulation reduced fitness. This calculation does not tolerate fitness estimates of zero, and so we excluded one glucocorticoid and one androgen study, each of which reported a fitness estimate of zero for one of their treatment groups. Results do not differ if we instead retained all data and calculated the scaled fitness ratio after adding 0.1 to each fitness estimate.

We conducted all analyses using R (version 3.5.2). We analyzed the glucocorticoid data using linear mixed effects (LME) models (function *lme* in the package *nlme*) with the relative fitness ratio (calculated as described above) as the response variable, and sex (male, female, both), life-history stage (breeding or non-breeding), hormone manipulation method (grouped into three categories: silastic, pellet/intraperitoneal/

osmotic, or oral/dermal), and fitness metric (reproductive success or survival) as fixed effects. We also included the interaction of sex by fitness metric, to test for the possibility that variation in the effect of hormone manipulations across the sexes depended on whether fitness was estimated using reproductive success or survival. We did not include age in the models because it was redundant with life history and fitness metric, given that juveniles are by definition not breeding and can only have survival assessed as a fitness estimate. Fourteen of the 19 estimates (74%) of effects of glucocorticoid augmentation on survival occurred in individuals of unknown sex or with data from males and females combined (i.e., sex = both), and 11 of those 19 studies (58%) assess effects in non-breeding individuals, so the effects of sex, life-history stage, and fitness metric are somewhat confounded in the dataset.

We considered two different random effects structures in our initial global models. One model included a random intercept for class (i.e., bird, fish, reptile, or mammal) to provide a coarse test for effects of phylogeny, as well as a random intercept for study identity, to account for inclusion of multiple fitness effects from the same study. We compared global models with class and study ID to models with only study ID using Aikaike's Information Criterion, corrected for small sample size (AICc), to select the random effect structure for all subsequent analyses (Supplementary material, Table S2). We did not consider models without study ID, as our inclusion of multiple fitness effects from individual studies would violate assumptions of independence of estimates. We weighted all analyses by the study's total sample size (control + treatment individuals, log-transformed to reduce undue influence of a few studies with exceptionally large sample sizes), based on the expectation that larger sample size studies might generate more accurate estimates of the fitness effects of hormone manipulations. Results are qualitatively unchanged if we analyze data without this weighting.

The relative fitness ratios for the glucocorticoid data set were non-normally distributed (as determined by Shapiro-Wilk test), and simple transformations did not improve the distribution, so we ran the analyses and identified five influential outliers (using the *plot* function with the specification "id = 0.05" in *nlme*). Removing these data points resulted in a normal distribution for the fitness ratios, so we report findings from analyses with these points excluded (resulting in an $N = 56$ observations from 36 studies). If we instead retained the outliers and transformed the fitness ratios using a complex modulus transformation (John and Draper, 1980), the distribution was improved, although still marginally non-normal. Analyses using the modulus-transformed data generate very similar results and identical conclusions to the analyses we present here with the outliers excluded. Excluding outliers was preferable because it allowed for comparison of fitness effects between the glucocorticoid and androgen analyses.

We analyzed androgen data using separate LMEs, because several of the factors that varied in the glucocorticoid dataset were less variable for the androgen dataset. For example, only three of the 26 studies employed pellet implants, with all of the remaining studies using silastic. All but one study was conducted in adults, and only three studies assessed effects in non-breeding individuals. As such, we did not include the fixed effects of manipulation method, age, or life-history stage in our analyses, but we did include these data in the analyses. Fitness effects estimated in the one study of juveniles and in three studies of non-breeding individuals fell well within the range reported in studies of breeding adults. Further, all studies considered effects for only one sex, or for the two sexes separately, so there are only two levels of sex for the androgen data (male or female), whereas there are three levels for the glucocorticoid data (both sexes combined, male, or female). The LME models for the androgen analyses included the relative fitness ratio as the response variable, and sex, fitness metric (reproductive success or survival), and their interaction as fixed effects. We assessed the random effects structure using the same approach as described above.

We simplified initial models using comparison of AICc values across recombinant candidate models with all possible combinations of fixed effects using the *dredge* function in the *MuMIn* package. We report results from the top-ranked model and contrast it with results from model averaging, if more than one model fell within 2 AICc of the top-ranked model. We checked fit of all global and final models by inspecting q-q plots, distributions of model residuals, and relationships between residuals and predictors. In the initial global LME model of the glucocorticoid data, model residuals varied across levels of life-history stage (breeding, non-breeding), and so we added a term to the LMEs to model the variance across levels of life-history stage to correct for this issue (using the *varIdent* function) (Pinheiro and Bates, 2006). The intercept of the null model provides a test of the Optimal Endocrine Phenotype Hypothesis, by testing whether the mean relative fitness ratio across all studies is greater or less than zero. A test of the variation in fitness ratios not explained by fixed effects (i.e., the intercept of the top model) provides an additional test of the hypothesis, by determining whether the relative fitness ratio remains different from zero after controlling for variation explained by fixed effects.

3. Results

We did not find evidence that the fitness effects of either glucocorticoid or androgen manipulations varied across taxa, as inclusion of taxonomic class as a random effect did not improve model fit relative to a global model that only included study ID as a random effect (AICc for models with only study ID > 3AICc lower than models with class + study ID). As such, all global models described below only included study ID as a random effect.

Overall, glucocorticoid augmentations reduced fitness (Fig. 2; intercept-only LME: $N = 56$, intercept = -0.15 ± 0.03 SE, $t = -4.82$, d.f. = 36, $p < 0.001$), even after controlling for fixed effects that predicted variation in the fitness ratio (top-ranked LME: $N = 56$, $F = 33.18$, d.f. = 1,35, $p < 0.001$). The top-ranked model to explain variation in fitness effects of glucocorticoids only retained the fixed effect of sex, but two other models were also ranked within 2 AICc (Supplementary material, Table S2). Model selection and model

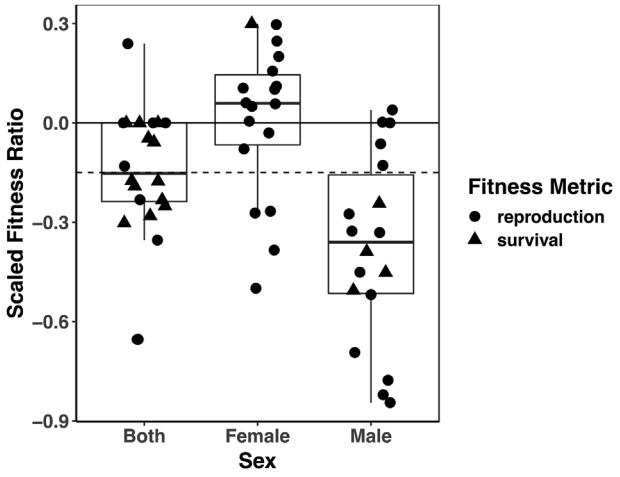

Fig. 2. Effects of exogenous glucocorticoids on relative fitness ($N = 56$, y-axis: LN-transformed fitness ratio of hormone-manipulated individuals relative to controls) varied depending on the sex of the individual being manipulated. On average, glucocorticoid manipulations reduce fitness (dashed line). Points represent individual fitness effects using data from original studies, and are jittered along the x-axis to facilitate visualization. Values below 0 (solid horizontal line) represent cases where the hormone manipulation reduced fitness. Effects of manipulations on fitness did not depend on the fitness metric used (circles = reproductive success, triangles = survival). Horizontal lines in each boxplot illustrate the median effect, boxes span from the 25th to 75th percentile of the data, and whiskers represent 1.5 inter-quartile ranges. See main text and Table S1 for statistical support.

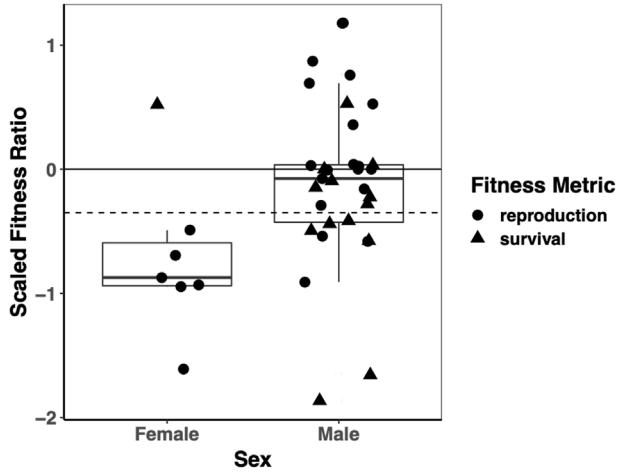

Fig. 3. Effects of exogenous androgens on relative fitness (N = 38, y-axis: LN-transformed fitness ratio of hormone-manipulated individuals relative to controls) varied with the sex of the individual being manipulated, depending on the fitness metric measured (circles = reproductive success, triangles = survival). On average, androgen manipulations reduced fitness (dashed line). Points represent individual fitness effects using data from original studies, and are jittered along the x-axis to facilitate visualization. Values below 0 (solid horizontal line) represent cases where the hormone manipulation reduced fitness. Horizontal lines in each boxplot illustrate the median effect, boxes span from the 25th to 75th percentile of the data, and whiskers represent 1.5 inter-quartile ranges. See main text for statistical support.

averaging did not provide evidence that fitness effects varied with method of hormone manipulation, life-history stage, treatment method, or fitness metric. Fitness was reduced most in studies involving assessment of males, and was more moderately reduced where both sexes were manipulated and assessed together, whereas, on average, fitness was not affected by glucocorticoid manipulations in females (Fig. 2; top-ranked LME: $N = 56$, sex $F = 11.85$, d.f. = 2,18, $p < 0.001$; post-hoc pairwise comparisons of least-squares means via Tukey test: both-female, -0.15 ± 0.07, $p = 0.091$; female-male, 0.40 ± 0.08, $p < 0.001$; both-male, 0.25 ± 0.07, $p = 0.005$). This average effect of glucocorticoid manipulations on the fitness of females reflects a wide range of variation across studies – in several cases (11 out of 18 studies, 61.1%) fitness increased in response to glucocorticoid manipulations, whereas in others, fitness was largely unaffected (1 of 18 studies, 5.6%) or reduced (6 of 18 studies, 33.3%).

Across all studies, androgen augmentations reduced fitness (Fig. 3; intercept-only LME: $N = 38$, intercept $= -0.35 \pm 0.13$ SE, $t = -2.73$, d.f. = 26, $p = 0.011$), even after controlling for fixed effects that predicted variation in the fitness ratio (top-ranked LME: $N = 38$, $F = 11.28$, d.f. = 1,24, $p = 0.003$). The average fitness effect of androgen manipulations was greater than the average effect of glucocorticoids on fitness (-0.35 vs -0.15, respectively, based on the intercept of the null LME model for each hormone). The full model was the top-ranked model to explain variation in fitness effects of androgen manipulations (AICc > 7 lower than the next-ranked model). The fitness effects of androgen manipulation varied between males and females in a way that depended on the fitness metric being measured (Fig. 3; global LME: $N = 38$; sex:fitness metric interaction, $F = 10.89$, d.f. = 1,10, $p = 0.008$; sex, $F = 3.27$, d.f. = 1,24, $p = 0.083$; fitness metric, $F = 7.65$, d.f. = 1,10, $p = 0.020$). In females, only one study assessed effects on survival, which was increased in response to androgen manipulation. The remaining six estimates in females documented reduced reproductive success in response to androgen manipulation. In contrast, in males, reproductive success appeared to be equally likely to be increased as decreased in response to androgen manipulation (9 increased, 6 decreased, 3 relatively unchanged), but survival was more consistently reduced (10 decreased, 2 increased, 1 unchanged).

4. Discussion

Overall, the results of our meta-analysis support the central prediction of the Optimal Endocrine Phenotype Hypothesis. Across studies, fitness was reduced by experimentally increasing either glucocorticoids (Fig. 2) or androgens (Fig. 3). However, the effects of experimentally elevated hormone levels on fitness also varied depending on sex of the individual and the component of fitness (survival or reproduction). Treatment with exogenous glucocorticoids consistently reduced both survival and reproductive success in males, whereas the effects of elevated glucocorticoids in females ranged from highly negative to highly positive, with no overall mean effect on fitness. Likewise, treatment with exogenous androgens consistently reduced reproductive success in females, whereas the effects of elevated androgens on reproductive success in males ranged from highly negative to highly positive, with no overall mean effect on fitness. While the reproductive success of males often increased in response to exogenous androgens, survival of males was more consistently reduced by experimentally increasing androgen levels.

The variation we found in the effects of hormone manipulations across the sexes is striking and suggests important differences in the ways that hormones regulate life history investment across males and females, which is in agreement with a large body of literature documenting sex differences in the organization and responsiveness of the glucocorticoid axis (i.e., HPA/I axis) (reviewed in Goel et al., 2011; Kudielka and Kirschbaum, 2005). Glucocorticoids can stimulate mobilization of resources and increased resource acquisition (Astheimer et al., 1991; Dallman et al., 1993; Pravosudov, 2003). For breeding females, who often play a larger role in offspring care (Queller, 1997), glucocorticoids might signal a need to upregulate resource investment into offspring (Crossin et al., 2012; Love et al., 2004). In males, where restraint in parental investment is more common, the same signal might indicate worsening conditions and a need to reduce investment in reproduction. However, glucocorticoid manipulations were also detrimental to survival in males, suggesting either a lack of increased investment in self-maintenance or a cost of dysregulation of the endocrine axis resulting from hormone manipulations. Comparative analyses of correlations between circulating glucocorticoids and life history strategies and environmental context provide some support for this idea (Bokony et al., 2009; Vitousek et al., 2019). Bokony et al. (2009) found that birds with higher brood value had higher baseline concentrations of glucocorticoids, but lower acute stress-induced concentrations. Several previous correlative analyses have also detected sex differences in glucocorticoid-fitness correlations (e.g., Bonier et al., 2006; Vitousek et al., 2018), which could reflect differential roles of glucocorticoids in regulating life history. Further comparative studies that contrast the effects of hormone manipulations across taxa and sexes with differing reproductive investment could clarify the potential for life history strategy-dependent effects of manipulations on life history investment, but we do not currently have sufficient data for such an analysis.

The available dataset provides a broad and synthetic test of the Optimal Endocrine Phenotype Hypothesis, but does suffer from some limitations. One such limitation is that fitness effects in the original studies are considered separately for individual fitness components. If hormones regulate a balance of tradeoffs in investment in competing life history traits (e.g., self-maintenance vs. reproduction), then a positive effect of a hormone manipulation on one fitness component could be negated by a counteracting, but unmeasured, effect on another component (Crossin et al., 2016; Ketterson et al., 1996; Love and Williams, 2008; Reed et al., 2006). For example, augmentation of circulating androgens could induce males to invest more in mate attraction but at a cost to the survival of their offspring, due to reductions in parental care (Clark and Galef, 1999; Ketterson and Nolan, 1994; Raouf et al., 1997). The across-studies patterns that we found are suggestive of

such counterbalancing, particularly for effects of androgens in males. However, integrated fitness estimates that incorporate both survival and reproductive success across a longer timespan (ideally approximating lifetime reproductive success) would be required for a full test of the Optimal Endocrine Phenotype Hypothesis.

A second important limitation of the dataset comes from the fact that many manipulations do not alter circulating hormones in a biologically relevant manner (Fusani, 2008). Silastic implants, the most commonly used method in our dataset, can induce supraphysiological spikes in circulating hormone, followed by suppression of circulating concentrations, sometimes below levels found in control animals (Goutte et al., 2011; Quispe et al., 2015). Both the Optimal Endocrine Phenotype and the Ongoing Selection Hypotheses make predictions about the effects of biologically relevant changes to the endocrine trait (Fig. 1). The drastic shifts in endocrine traits resulting from many hormone augmentation methods could incur larger fitness costs. In some regards, this might mean the data provide a strong test of the Optimal Endocrine Phenotype Hypothesis, because the differences between treatment and control groups are large enough to ensure a treatment effect can be detected, even in a small sample size experiment. However, we cannot conclude that more biologically relevant manipulations would similarly affect fitness. We also cannot distinguish whether the observed fitness effects of hormone augmentations are due to altered circulating concentrations or due to changes in the responsiveness and capacity of the endocrine axis. The glucocorticoid dataset included studies that employed an array of manipulation methods, including some that should only transiently and moderately influence circulating hormone concentrations (e.g., oral dosing). This variation in manipulation technique did not predict variation in the fitness effect of the manipulation. As more primary research documenting fitness effects of conservative hormone manipulations accumulates, we will be better able to determine the degree to which the mode of manipulation influences the magnitude or even direction of the fitness effect.

We could not directly test the Ongoing Selection Hypothesis because it requires that studies deliberately shift individuals with putatively maladaptive endocrine phenotypes toward a hypothesized optimal trait value. To our knowledge, such an experiment has not been conducted to date. Logistically, such a test would require an initial collection of data from which the optimal trait value could be inferred (e.g., using the population level trait-fitness relationship to infer a fitness surface), and then screening of individuals to be included in the hormone manipulation experiment to determine which would be predicted to benefit from hormone augmentation (i.e., which individuals had a putatively suboptimal phenotype). Manipulations could be conducted only in these individuals, or, alternatively, across all individuals, but with predictions for fitness effects differing depending on the starting endocrine phenotype of each individual. This latter approach should be feasible in most field contexts, where tissue samples collected at the time of hormone manipulation treatments could provide information on the pre-manipulation endocrine phenotype, which can then be used to determine if the fitness effects depended on an individual's pre-manipulation distance from the hypothesized optimal trait value. Under the Ongoing Selection Hypothesis, fitness costs would be expected for individuals that started with a near-optimal phenotype and fitness benefits would be expected for those further from the optimum. In contrast, in this same experiment, fitness costs would be predicted across all individuals under the Optimal Endocrine Phenotype Hypothesis. Such concurrent experimental tests of these two competing hypotheses would help advance our understanding of the adaptive significance of individual variation in endocrine phenotypes.

While these two hypotheses generate distinct predictions regarding the effects of hormone manipulations, they represent opposite ends of a continuum in explaining individual variation in endocrine phenotypes, and both could be supported in different scenarios, even within the same population. The crucial difference between the two hypotheses is the degree to which past responses to selection and current plasticity (or flexibility) in endocrine phenotypes are sufficient to allow individuals to track optima that change as the environment changes. Given the dynamic nature of endocrine phenotypes and many other physiological phenotypes, we might expect that they are better able to shift adaptively and match environmentally dictated optima than many other traits. However, in some individuals or some environments, optimal expression of endocrine phenotypes might be more or less possible, tipping the balance toward or away from the Optimal Endocrine Phenotype Hypothesis. This variation could be driven by differences among individuals, such as developmental constraints that could influence an individual's ability to express optimal endocrine phenotypes as an adult (Farrell et al., 2015). We could also see variation across environments, with rapid or novel changes beyond a population's historic norms (e.g., due to climate change or urbanization) being more difficult for individuals to track through phenotypic plasticity (Chevin et al., 2010; Sih et al., 2011; Wingfield, 2013). Over time, adaptive plasticity in populations that are able to persist in such environments would be expected to evolve, so that individuals are better able to express phenotypes that match the challenges they face (Ghalambor et al., 2007; Price et al., 2003). This generates an interesting and testable prediction – individuals in newly-established introduced populations ought to more often express maladaptive plastic phenotypes, and more closely follow predictions of the Ongoing Selection Hypothesis. After several generations, populations introduced into novel environments might provide increasing support for the Optimal Endocrine Phenotype Hypothesis over time, as they evolve in response to the novel selective pressures they face in their new habitat (Agrawal, 2001; Richards et al., 2006; Sexton et al., 2002).

The field of evolutionary endocrinology continues to make important advances, as an increasing number of biologists consider individual variation in endocrine phenotypes through an evolutionary lens. We hope that future deliberate tests of the Optimal Endocrine Phenotype and Ongoing Selection Hypotheses will provide new appreciation for the scope of adaptive plasticity and the degree to which natural selection continues to shape endocrine traits.

Acknowledgments

We thank Paul Martin, Sean Lema, and two anonymous reviewers for insightful feedback on earlier versions of this manuscript, as well as Ignacio Moore, Laura Schoenle, Jenny Ouyang, Adam Lendvai, Michaela Hau, Creagh Breuner, and Joel McGlothlin for discussions that have shaped the ideas we present. We are also grateful to several researchers who provided data for use in our analyses (listed in Table S1).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mce.2019.110640.

References

Agrawal, A.A., 2001. Phenotypic plasticity in the interactions and evolution of species. Science 294, 321–326.
Arnold, S.J., Pfrender, M.E., Jones, A.G., 2001. The adaptive landscape as a conceptual bridge between micro-and macroevolution. Genetica 112–113, 9–32.
Astheimer, L.B., Buttemer, W.A., Wingfield, J.C., 1991. Interactions of corticosterone with feeding, activity and metabolism in Passerine birds. Ornis Scand. 23, 355–365.
Bokony, V., Lendvai, A.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.
Bonier, F., Martin, P.R., 2016. How can we estimate natural selection on endocrine traits? Lessons from evolutionary biology. Proc. R. Soc. Biol. Sci. 283 20161887.
Bonier, F., Martin, P.R., Moore, I.T., Wingfield, J.C., 2009a. Do baseline glucocorticoids predict fitness? Trends Ecol. Evol. 24, 634–642.
Bonier, F., Martin, P.R., Sheldon, K.S., Jensen, J.P., Foltz, S.L., Wingfield, J.C., 2006. Sex-specific consequences of life in the city. Behav. Ecol. 18, 121–129.
Bonier, F., Moore, I.T., Martin, P.R., Robertson, R.J., 2009b. The relationship between fitness and baseline glucocorticoids in a passerine bird. Gen. Comp. Endocrinol. 163, 208–213.

Calsbeek, R., Gosden, T.P., Kuchta, S.R., Svensson, E.I., 2012. Fluctuating selection and dynamic adaptive landscapes. In: Svensson, E.I., Calsbeek, R. (Eds.), The Adaptive Landscape in Evolutionary Biology. Oxford University Press, pp. 89–109.
Chevin, L.-M., Lande, R., Mace, G.M., 2010. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. PLoS Biol. 8, e1000357.
Clark, M.M., Galef Jr., B.G., 1999. A testosterone-mediated trade-off between parental and sexual effort in male Mongolian gerbils (*Meriones unguiculatus*). J. Comp. Psychol. 113, 388.
Crossin, G.T., Love, O.P., Cooke, S.J., Williams, T.D., 2016. Glucocorticoid manipulations in free-living animals: considerations of dose delivery, life-history context and reproductive state. Funct. Ecol. 30, 116–125.
Crossin, G.T., Trathan, P.N., Phillips, R.A., Gorman, K.B., Dawson, A., Sakamoto, K.Q., Williams, T.D., 2012. Corticosterone predicts foraging behavior and parental care in Macaroni Penguins. Am. Nat. 180, E31–E41.
Dallman, M.F., Strack, A.M., Akana, S.F., Bradbury, M.J., Hanson, E.S., Scribner, K.A., Smith, M., 1993. Feast and famine: critical role of glucocorticoids with insulin in daily energy flow. Front. Neuroendocrinol. 14, 303–347.
Dey, C.J., O'Connor, C.M., Gilmour, K.M., Van Der Kraak, G., Cooke, S.J., 2010. Behavioral and physiological responses of a wild teleost fish to cortisol and androgen manipulation during parental care. Horm. Behav. 58, 599–605.
Farrell, T., Kriengwatana, B., MacDougall-Shackleton, S.A., 2015. Developmental stress and correlated cognitive traits in songbirds. Comp. Cogn. Behav. Rev. 10, 1–23.
Fusani, L., 2008. Endocrinology in field studies: problems and solutions for the experimental design. Gen. Comp. Endocrinol. 157, 249–253.
Ghalambor, C.K., McKay, J.K., Carroll, S.P., Reznick, D.N., 2007. Adaptive versus non--adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. Funct. Ecol. 21, 394–407.
Goel, N., Workman, J.L., Lee, T.T., Innala, L., Viau, V., 2011. Sex differences in the HPA axis. Compr. Physiol. 4, 1121–1155.
Goutte, A., Clément-Chastel, C., Moe, B., Bech, C., Gabrielsen, G.W., Chastel, O., 2011. Experimentally reduced corticosterone release promotes early breeding in black-legged kittiwakes. J. Exp. Biol. 214, 2005–2013.
Hau, M., Casagrande, S., Ouyang, J.Q., Baugh, A.T., 2016. Glucocorticoid-mediated phenotypes in vertebrates: multilevel variation and evolution. In: Naguib, M. (Ed.), Advances in the Study of Behavior. Elsevier, pp. 41–115.
Honarmand, M., Goymann, W., Naguib, M., 2010. Stressful dieting: nutritional conditions but not compensatory growth elevate corticosterone levels in zebra finch nestlings and fledglings. PLoS One 5, e12930.
John, J.A., Draper, N.R., 1980. An alternative family of transformations. J. R. Stat. Soc.: Ser. C Appl. Stat. 29, 190–197.
Ketterson, E.D., Nolan Jr., V., 1994. Male parental behavior in birds. Annu. Rev. Ecol. Systemat. 25, 601–628.
Ketterson, E.D., Nolan Jr., V., 1999. Adaptation, exaptation, and constraint: a hormonal perspective. Am. Nat. 154, S4–S25.
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.
Kudielka, B.M., Kirschbaum, C., 2005. Sex differences in HPA axis responses to stress: a review. Biol. Psychol. 69, 113–132.
Lande, R., Arnold, S.J., 1983. The measurement of selection on correlated characters. Evolution 37, 1210–1226.
Lendvai, Á.Z., Ouyang, J.Q., Schoenle, L.A., Fasanello, V., Haussmann, M.F., Bonier, F., Moore, I.T., 2014. Experimental food restriction reveals individual differences in corticosterone reaction norms with no oxidative costs. PLoS One 9, e110564.
Love, O.P., Breuner, C.W., Vézina, F., Williams, T.D., 2004. Mediation of a corticosterone-induced reproductive conflict. Horm. Behav. 46, 59–65.
Love, O.P., Williams, T.D., 2008. The adaptive value of stress-induced phenotypes: effects of maternally derived corticosterone on sex-biased investment, cost of reproduction, and maternal fitness. Am. Nat. 172, E135–E149.
Lynn, S.E., Breuner, C.W., Wingfield, J.C., 2003. Short-term fasting affects locomotor activity, corticosterone, and corticosterone binding globulin in a migratory songbird. Horm. Behav. 43, 150–157.
Martin, T.E., 1987. Food as a limit on breeding birds: a life-history perspective. Annu. Rev. Ecol. Systemat. 18, 453–487.
McGlothlin, J.W., Jawor, J.M., Ketterson, E.D., 2007. Natural variation in a testosterone-mediated trade-off between mating effort and parental effort. Am. Nat. 170, 864–875.
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.
Mitchell-Olds, T., Shaw, R.G., 1987. Regression analysis of natural selection: statistical inference and biological interpretation. Evolution 41, 1149–1161.
Patterson, S., Hahn, T., Cornelius, J., Breuner, C., 2014. Natural selection and gluco-corticoid physiology. J. Evol. Biol. 27, 259–274.
Pinheiro, J.C., Bates, D.M., 2006. Mixed-effects Models in S and S-PLUS. Springer, New York.
Pravosudov, V.V., 2003. Long-term moderate elevation of corticosterone facilitates avian food-caching behaviour and enhances spatial memory. Proc. R. Soc. Lond. Ser. B Biol. Sci. 270, 2599–2604.
Price, T., Kirkpatrick, M., Arnold, S.J., 1988. Directional selection and the evolution of breeding date in birds. Science 240, 798–799.
Price, T.D., Qvarnström, A., Irwin, D.E., 2003. The role of phenotypic plasticity in driving genetic evolution. Proc. R. Soc. Lond. Ser. B Biol. Sci. 270, 1433–1440.
Queller, D.C., 1997. Why do females care more than males? Proc. R. Soc. Lond. Ser. B Biol. Sci. 264, 1555–1557.
Quispe, R., Trappschuh, M., Gahr, M., Goymann, W., 2015. Towards more physiological manipulations of hormones in field studies: comparing the release dynamics of three kinds of testosterone implants, silastic tubing, time-release pellets and beeswax. Gen. Comp. Endocrinol. 212, 100–105.
Råberg, L., Grahn, M., Hasselquist, D., Svensson, E., 1998. On the adaptive significance of stress-induced immunosuppression. Proc. R. Soc. Lond. Ser. B Biol. Sci. 265, 1637–1641.
Raouf, S.A., Parker, P.G., Ketterson, E.D., Nolan Jr., V., Ziegenfus, C., 1997. Testosterone affects reproductive success by influencing extra–pair fertilizations in male dark–eyed juncos (Aves: *Junco hyemalis*). Proc. R. Soc. Lond. Ser. B Biol. Sci. 264, 1599–1603.
Reed, W., Clark, M., Parker, P., Raouf, S., Arguedas, N., Monk, D., Snajdr, E., Nolan Jr., V., Ketterson, E., 2006. Physiological effects on demography: a long-term experimental study of testosterone's effects on fitness. Am. Nat. 167, 667–683.
Richards, C.L., Bossdorf, O., Muth, N.Z., Gurevitch, J., Pigliucci, M., 2006. Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. Ecol. Lett. 9, 981–993.
Schluter, D., 1988. Estimating the form of natural selection on a quantitative trait. Evolution 42, 849–861.
Sexton, J.P., McKay, J.K., Sala, A., 2002. Plasticity and genetic diversity may allow saltcedar to invade cold climates in North America. Ecol. Appl. 12, 1652–1660.
Sih, A., Ferrari, M.C.O., Harris, D.J., 2011. Evolution and behavioural responses to human-induced rapid environmental change. Evol. Appl. 4, 367–387.
Sopinka, N.M., Patterson, L.D., Redfern, J.C., Pleizier, N.K., Belanger, C.B., Midwood, J.D., Crossin, G.T., Cooke, S.J., 2015. Manipulating glucocorticoids in wild animals: basic and applied perspectives. Conserv. Physiol. 3, cov031.
Taff, C.C., Vitousek, M.N., 2016. Endocrine flexibility: optimizing phenotypes in a dynamic world? Trends Ecol. Evol. 31, 476–488.
Tschirren, B., Postma, E., Gustafsson, L., Groothuis, T.G., Doligez, B., 2014. Natural selection acts in opposite ways on correlated hormonal mediators of prenatal maternal effects in a wild bird population. Ecol. Lett. 17, 1310–1315.
Vitousek, M.N., Johnson, M.A., Downs, C.J., Miller, E.T., Martin, L.B., Francis, C.D., Donald, J.W., Fuxjager, M.J., Goymann, W., Hau, M., 2019. Macroevolutionary patterning in glucocorticoids suggests different selective pressures shape baseline and stress-induced levels. Am. Nat. 193, 866–880.
Vitousek, M.N., Taff, C.C., Hallinger, K.K., Zimmer, C., Winkler, D.W., 2018. Hormones and fitness: evidence for trade-offs in glucocorticoid regulation across contexts. Front. Ecol. Evol. 6, 42.
Williams, T.D., 2007. Individual variation in endocrine systems: moving beyond the 'tyranny of the Golden Mean'. Philos. Trans. R. Soc. Biol. Sci. 363, 1687–1698.
Wingfield, J.C., 2013. The comparative biology of environmental stress: behavioural endocrinology and variation in ability to cope with novel, changing environments. Anim. Behav. 85, 1127–1133.