Integrative and Comparative Biology

Integrative and Comparative Biology, volume 58, number 4, pp. 751–762
doi:10.1093/icb/icy062

Society for Integrative and Comparative Biology

SYMPOSIUM

Standing Variation and the Capacity for Change: Are Endocrine Phenotypes More Variable Than Other Traits?

Meredith C. Miles,* Maren N. Vitousek,†,‡ Jerry F. Husak,§ Michele A. Johnson,¶ Lynn B. Martin,‖ Conor C. Taff,† Cedric Zimmer,† Matthew B. Lovern# and Matthew J. Fuxjager[1],*

*Department of Biology, Wake Forest University, Winston-Salem, NC 27109, USA; †Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA; ‡Cornell Lab of Ornithology, Ithaca, NY 14850, USA; §Department of Biology, University of St. Thomas, St. Paul, MN 55105, USA; ¶Department of Biology, Trinity University, San Antonio, TX 78212, USA; ‖Department of Global Health, University of South Florida, Tampa, FL 33620, USA; #Department of Zoology, Oklahoma State University, Stillwater, OK 74078, USA

From the symposium "Understanding the Evolution of Endocrine System Variation through Large-scale Comparative Analyses" presented at the annual meeting of the Society for Integrative and Comparative Biology, January 3–7, 2018 at San Francisco, California.

[1]E-mail: mfoxhunter@gmail.com

Synopsis Circulating steroid hormone levels exhibit high variation both within and between individuals, leading some to hypothesize that these phenotypes are more variable than other morphological, physiological, and behavioral traits. This should have profound implications for the evolution of steroid signaling systems, but few studies have examined how endocrine variation compares to that of other traits or differs among populations. Here we provide such an analysis by first exploring how variation in three measures of corticosterone (CORT)—baseline, stress-induced, and post-dexamethasone injection—compares to variation in key traits characterizing morphology (wing length, mass), physiology (reactive oxygen metabolite concentration [d-ROMs] and antioxidant capacity), and behavior (provisioning rate) in two populations of tree swallow (*Tachycineta bicolor*). After controlling for measurement precision and within-individual variation, we found that only post-dex CORT was more variable than all other traits. Both baseline and stress-induced CORT exhibit higher variation than antioxidant capacity and provisioning rate, but not oxidative metabolite levels or wing length. Variation in post-dex CORT and d-ROMs was also elevated in the higher-latitude population in that inhabits a less predictable environment. We next studied how these patterns might play out on a macroevolutionary scale, assessing patterns of variation in baseline testosterone (T) and multiple non-endocrine traits (body length, mass, social display rate, and locomotion rate) across 17 species of *Anolis* lizards. At the macroevolutionary level, we found that circulating T levels and the rate of social display output are higher than other behavioral and morphological traits. Altogether, our results support the idea that within-population variability in steroid levels is substantial, but not exceptionally higher than many other traits that define animal phenotypes. As such, circulating steroid levels in free-living animals should be considered traits that exhibit similar levels of variability from individual to individual in a population.

Introduction

A major goal of evolutionary endocrinology is to evaluate how hormone systems change over time to support behavioral adaptation and ultimately diversification (Crews and Moore 1986; Hau 2007; Adkins-Regan 2008; Williams 2008; Wingfield et al. 2008; Ketterson et al. 2009; McGlothlin et al. 2010). Studies that explore this topic often assume that endocrine traits exhibit heritable variation and are thus evolvable. Evidence certainly supports this idea (King et al. 2004; Jenkins et al. 2014; Stedman et al. 2017), and thus lends credence to the notion that endocrine systems may change over time in response to selection. One important characteristic of endocrine phenotypes is that they often exhibit high standing variation within populations (i.e., the tendency for individuals to deviate from the mean), but it is unknown how this determines the capacity for an

Advance Access publication June 12, 2018

entire trait distribution to shift in response to processes such as selection and drift at the population/species level (Kempenaers et al. 2008; Williams 2008). Similarly, we know little about how endocrine variability compares with other traits that define an animal's morphology, behavior, and non-endocrine physiology.

Here we attempt to address this gap by quantifying and comparing population-level variation in morphological, behavioral, and physiological (including endocrine) traits through two approaches: (i) by comparing variance between different traits present within a single population, and (ii) seeing how variable each trait is by comparing its variance among different populations and species. Although most studies adopt an overt focus on measures of central tendency to describe phenotypes, this overlooks the fact that phenotypic variation itself should also be the product of evolutionary processes including selection and drift (Bennett 1987; Williams 2008). This is because higher population-level variation reflects a larger proportion of individuals with phenotypes outside the norm, which may experience differential fitness outcomes in the face of sudden directional selection (e.g., Grant 1999). As an illustrative example, consider how two populations with identical means for a given phenotype—but dramatically different variances—would each be impacted by identical directional selection. Individuals farther away from the selection optimum will not survive and reproduce, and there are more individuals near the optimum in the highly variable population. As a result, the population with higher standing variation will undergo a shift in the mean toward the selection optimum while the phenotypically inflexible population will be at a higher risk for extinction (Freeman and Herron 2007). Of course, we can also consider this process through a historical lens; just as present-day phenotypes may be subject to future selection, they are also the product of past evolutionary processes. Investigating how these processes operate on populations with different trait variances—instead of differences in the mean alone—is therefore necessary to fully understand how traits evolve and impact future fitness. This framework is seldom integrated into comparative endocrinology, which is surprising, considering the outsized role played by hormones in governing an animal's capacity to respond to its environment (Williams 2008).

Biologists have long noted that steroid levels can vary within individuals, rapidly changing in response to stimuli from its physical and social environment (Wingfield et al. 1990; Wingfield and Sapolsky 2003; Hirschenhauser and Oliveira 2006; Hau 2007; Kempenaers et al. 2008; Williams 2008; Gleason et al. 2009; Hau et al. 2010; Breuner et al. 2013; Taff and Vitousek 2016). Because of this flexibility, one might hypothesize that standing variation in steroid levels among individuals within a given population is similarly high. To this end, we might even suspect that such population-level variability is higher than that of many other traits that define an animal's phenotype. However, levels of variability in circulating steroids are seldom quantified and compared with other traits, despite numerous calls for such analyses within the field of evolutionary endocrinology (Kempenaers et al. 2008; Williams 2008; Taff and Vitousek 2016).

Two widely studied steroid hormones are corticosterone (CORT) and testosterone (T). CORT is a glucocorticoid hormone that binds to both mineralocorticoid and glucocorticoid receptors, mediating a variety of metabolic functions (Wingfield and Sapolsky 2003). In particular, CORT modulates metabolic processes and helps an organism physiologically and behaviorally respond to stressors in their environment, while also facilitating the animal's return to a homeostatic baseline following such experiences (Sapolsky et al. 2000). Likewise, T is an androgenic steroid that binds to either androgen receptors (before or after conversion to dihydrotestosterone) or estrogen receptors (after conversion to estradiol). This hormone is secreted by the gonads and plays an important role in the organization and activation of the masculine reproductive phenotype (Wingfield et al. 2000). Even though organismal endocrinologists have studied the evolution and function of these two steroid systems for decades, we still know little about their variability in free-living animals.

At the population level, we assessed variability in CORT across two populations of tree swallows (*Tachycineta bicolor*), a geographically widespread bird species. The two populations (one in New York and one in Alaska) provide an excellent opportunity to understand how variability changes among well-diverged populations (or incipient species) and across distinct selection regimes. Because one population is restricted to the subarctic biogeographic zone, it is subject to less predictable ambient conditions and a compressed breeding season relative to its temperate counterpart. The highly variable conditions present in the arctic may therefore lead to fluctuating selection, which should result in higher standing variation in multiple traits. As such, we compared standing variation in CORT to other traits, spanning morphology (body mass and wing length), physiology (measures of oxidative damage [d-ROMs] and plasma antioxidant capacity [OXY]), and behavior (nest provisioning rates). Current evidence suggests that all of these traits can exhibit at least some heritable variation, and thus can evolve in response to evolutionary processes such as selection and drift (Tsuji et al. 1989; MacColl and Hatchwell 2003; Costantini and Dell'Omo 2006; Olsson et al. 2008; Losdat et al. 2014).

If variability itself evolves in response to selection and/or drift, then there should be distinct differences in variability between traits on a macroevolutionary scale as well. This is impossible to study in only two recently-diverged populations, of course, so we next examined how within-population variation in circulating T levels compared with variation in other traits in a radiation of tropical lizards (*Anolis* sp.). This genus is ideal for such an analysis, as *Anolis* sp. are geographically widespread, morphologically and behaviorally diverse, and have a well-resolved phylogeny. We hypothesized that population-level variance should evolve differently when comparing circulating hormone levels with morphological (body mass and snout-to-vent length) and behavioral traits (frequency of locomotion and social displays, including aggressive and courtship displays). Androgenic systems are thought to play a major role in activating sexual behavior and perhaps driving its evolution (Fuxjager et al. 2018), with particularly strong influence on vigorous courtship displays and reproduction (Holmes and Wade 2005; Holmes et al. 2007; Johnson and Wade 2010; Husak and Lovern 2014; Fuxjager et al. 2017; Johnson et al. 2018). When it comes to these sexual behaviors specifically, previous work also suggests that they are more evolutionarily labile than morphological and physiological variables not related to sexual reproduction (Blomberg et al. 2003). Therefore, endocrine variability may be similarly variable to sexual behavior to facilitate its evolution.

Methods

Tree swallows

Tree swallows sampled for this study were from populations breeding in Ithaca, NY, USA ($n = 148$ for physiological/morphological data and $n = 19$ for behavioral data) and McCarthy, AK, USA ($n = 63$, physiological/morphological data only). Morphological and physiological data were collected in 2016 (NY: May 17–July 1, AK: June 2–July 7) by capturing individuals during their respective breeding seasons in nest boxes, either by hand or by way of a manually activated trap door installed on the box. The following procedures were approved by all appropriate institutional and governmental authorities.

Upon each individual's capture, we collected blood from the brachial wing vein within 3 min of capture following protocols described in detail elsewhere (Stedman et al. 2017). Stress-induced CORT was measured using a standardized restraint stress protocol (Cockrem 2013), whereby a second blood sample was collected 30 min following an initial disturbance (Stedman et al. 2017). A synthetic glucocorticoid was then injected (dexamethasone [dex]: 1.5 μg/g) to test the ability to terminate the stress response through negative feedback; a final blood sample was collected 30 min post injection. Blood was collected between 0700 and 1000 to minimize variation due to circadian rhythms. Blood samples from females were collected on days 6–7 of incubation; males were sampled on days 3–7 of the nestling provisioning period (Vitousek et al. 2018). At the same time, we also measured each individual's body mass using a Pesola spring balance (to the nearest 0.25 g), and flat wing length (to the nearest 0.5 mm). We intentionally collected repeated morphological measures of some individuals ($n = 82$), allowing us to compute within-individual coefficients of variation (CV) for both mass (average population CV =7.5%) and wing length (1.0%). We suspect that mass CV appears to be higher because body mass (like hormone levels) fluctuates within an individual from day to day and across the annual cycle (Dunning 1992). As such, whereas within-individual CV for wing length reflects measurement precision alone, the CV for mass reflects the combined influence of measurement precision and actual biological variation within an individual.

Following collection, blood samples were kept on ice until centrifugation, and frozen at −30°C until analysis. Steroids were isolated from plasma using a triple ethyl acetate extraction (detailed previously in Stedman et al. 2017), after which CORT levels were assayed in duplicate alongside a nine-point standard curve using a commercially available EIA Kit (Detect Corticosterone, Arbor Assays; K104-H5). Validation tests confirmed that following extraction, this procedure showed parallelism in tree swallow plasma (detailed in Taff et al. 2018). Samples were extracted using a starting volume of 5 μL of plasma; for the few samples in which <5 μL of plasma was available, we utilized the maximum possible starting volume and corrected for the dilution. Extraction efficiency averaged 85.4%. Samples from both populations were run across multiple plates; inter-assay variation was 5.7% and intra-assay variation was 5.4%.

Using baseline blood samples, we also assessed oxidative damage and antioxidant capacity using d-ROMs kits and OXY-adsorbent tests, respectively (Vitousek et al. 2016, in revision). The d-ROMs kit (Diacron International, Grosseto, Italy) quantifies the concentration of reactive oxygen metabolites that result from the oxidation of biomolecules, whereas the OXY-adsorbent test (Diacron International) assesses the ability of plasma antioxidants to resist antioxidants. All samples were run in duplicate according to the kit manufacturer's instructions. For the d-ROM assay, intra-plate variation was 10.9% and inter-plate variation was 6.3%. For the OXY-adsorbent test, intra-plate variation was 10.1% and inter-plate variation was 12.0%.

For the New York population only, we also monitored nestling provisioning behavior (feeds per day) in a separate group of individuals (May–July 2015; Vitousek et al. in revision). This allowed us to explore how variability in behavior compared with that of CORT levels, even if we could not compare such variability across populations. Briefly, females and their mates were fitted with radio frequency identification (RFID) tags (2× 12 mm; EM4102, Cyntag Inc., Cynthiana, KY, USA). Every time a bird passed through a copper-wire antenna fitted around the entrance to their nest box, their RFID tag's unique identification code was recorded along with a timestamp. RFID reader boards (Bridge and Bonter 2011) were set to record during all daylight hours (0500–2200) from hatching day to 18 days after hatching (when tree swallows begin to fledge). Poll time was set at 0500, and cycle time at 1000, with a delay of 1. Raw read sequences were used to estimate daily nestling feeding rates using an algorithm designed to identify distinct feeding visits (described in detail in Vitousek et al. in revision). From these records we obtained total daily feeding rates for each individual from 20 individuals.

Tree swallow data analysis

We used a model comparison approach to see if trait variation was different between different traits of a given population, or otherwise different between both populations for each of our measured traits (CORT levels, dROMs, OXY, provisioning rate, wing length, and mass). This is because we could not use a statistical approach that relied on direct comparisons of CV (CV $= \frac{\sigma}{\bar{x}}$), as each population only had a single CV for a given trait (morphological, physiological, and endocrine variables described above). To compare variability among traits, we first accounted for differences in scaling among variables. We therefore performed a standard rescaling for each distribution to have minimum $=0$ and maximum $=100$, which fits the distribution of every variable into the 0–100 range without changing the proportional relationships around the mean and variance. This provides a basis for comparisons while avoiding the confounding effect of differences in measurement units among variables.

Next, we examined variability in traits between two populations of tree swallows (from New York and Alaska). Therefore, we instead adopted a model-comparison approach by constructing linear mixed models (LMMs) with population identity as the single fixed factor on a given trait (the response) and sex as a random factor. For each trait, we then fit two similar models: (i) a typical LMM where both populations are assumed to have equal variance, and (ii) a more complex model ($\mathrm{df}_{\mathrm{alternative}} = \mathrm{df}_{\mathrm{null}} + 1$) where each population could have its own variance. By comparing the fit of these nested models with the likelihood ratio test (LRT), we tested the statistical hypothesis that the data are better explained by a model where populations exhibit differences in phenotypic variability. Therefore, a significant ($P < 0.05$ after correction for multiple testing) LRT result supports the inference that a given trait is differentially variable between populations.

We also used this approach to compare endocrine CORT variability (baseline, stress-induced, and post-dex) with other traits, with the goal of evaluating whether circulating hormone levels are hyper-variable traits. In this case, we were no longer examining population differences in variability, but instead were examining how variability compares among traits themselves. The overall modeling approach was the same as described above (i.e., fit two LMMs—one with equal variances and one with unequal variances, followed by a comparison to evaluate whether distribution variabilities are unequal). We included sex (nested in population) as a random factor. To account for potential difficulties when comparing very different traits within a population, we also included a measure of precision as a covariate. For morphology and behavior, where we had more than three observations per individual, we were able to compute a within-individual CV to characterize the precision of both measurements. However, note that avian body mass is known to fluctuate on day-to-day scales (Dunning 1992), which means that the CV for mass also likely reflects genuine within-individual variation (though it is impossible to distinguish "genuine" variation from precision in this case). Because only some individuals had multiple measurements available, we also restricted these analyses to individuals for which we had repeated morphology measures. For physiological assays, we did not have multiple independent samples from single individuals and samples were only run in duplicate, so we could only account for measurement precision using the intra-plate CV. There was no relationship between precision/error/within-individual CV ($R^2 = 0.10$, $P = 0.487$), which suggests that this did not confound our analyses. After fitting the entire model series and evaluating whether equal or unequal variances better fit the data for each trait pair, we again adjusted all *P*-values to control the false discovery rate.

Anole data collection

Data for the *Anolis* lizards used herein (see Supplementary Table S1 for list of species) are described in detail by Husak and Lovern (2014). Briefly, we sampled species at Discovery Bay Marine Laboratory, Jamaica; in and around the "Fountain of Youth" and the Bimini Nature Trail on South Bimini, Bahamas; in and around Coralsol Resort near Barahona, Dominican Republic; and in El Yunque National Forest in and around El Verde Field Station, Puerto Rico. Samples were collected over 2 years, but in small a window of time (2 weeks per site) within the same year and always during the breeding season. At each site, we sampled between 0800 and 1200.

For each captured individual, we collected blood from the suborbital sinus with a heparinized microhematocrit tube. Blood sampling occurred within 4 min of capture to avoid elevation of CORT and/or possible effects on T levels (Langkilde and Shine 2006; Baird and Hews 2007). We kept blood on ice packs until it was centrifuged to separate the plasma, which was then frozen until later processing in the USA. We measured plasma T via radioimmunoassay (RIA) after extracting and isolating steroids with column chromatography (see Husak and Lovern 2014). Samples were adjusted for individual recovery and initial sample volume, with an average T recovery of 77%, intra-assay CV of 12.0%, and inter-assay CV of 6.2%. Assay sensitivity was <10 pg/mL. After blood samples were taken, we measured snout–vent length (SVL) to the nearest 0.1 mm with digital calipers and mass with a spring scale to the nearest 0.1 g.

Behavioral data for 14 of these 16 anole lizard species were collected in the same populations (but not the same individuals, and during different years). Note, however, that we did not collect behavioral data from *A. garmani*, whereas we collected behavioral data from *A. distichus* near Bani, Dominican Republic (hormone data for this species were collected in Bimini, Bahamas). Subsets of these behavioral data have been previously published (Johnson and Wade 2010; Johnson et al. 2018). To quantify locomotor and pushup behaviors, we observed adult male lizards of each species during the 2004–2015 summer (May–July) breeding seasons. All observations occurred between 0700 and 1800, and never during inclement weather (i.e., rain). Individual lizards were located for observation by slowly walking through the field sites, and when an undisturbed lizard was identified, it was observed for 10–120 min from a minimum distance of 10 m. During observations, we recorded all locomotor behaviors (with each movement defined as a crawl, run, or jump) and pushup displays (noting each up-and-down movement as a single pushup), to determine the rates of total locomotor movements and pushups for each observation (details in Supplementary Table S1).

Anole data analysis

In a final analysis, we compared variability in traits among 17 *Anolis* species (Supplementary Table S1). This analysis was different (and more straightforward) than the ones above, because direct comparison of variability across many species is feasible using CV with a larger comparative dataset. Because we were comparing paired traits to one another within a set group of species, ran a phylogenetic paired *t*-test in the R package "phytools" (Revell 2012; a general package for comparative methods) to conduct pairwise comparisons of each trait's normalized CV (i.e., CV computed from a distribution that has been rescaled to fit range 0–100, described above) to each other. This is analogous to running Tukey post hoc comparisons in an analysis of variance, so we conservatively Holm-adjusted all output *P*-values from this analysis. We also tested whether variation was correlated among traits using pairwise phylogenetic generalized least squares (PGLS; analogous to linear regression). Again, all *P*-values reported were corrected for multiple testing.

Results

Variability at the population level

We modeled the influence of population (New York or Alaska) on the variability of each trait by comparing the fit of models in which they were either permitted to have unique variances or the variances were equal between populations (Table 1). If the more complex, unequal-variances model provided a

Table 1 Fit comparison among models testing for an effect of population (Alaska vs. New York) on tree swallow trait distributions, where the null model constrained variances to remain equal and the more complex (alternative) model allows for unequal variances among the populations

Trait	Model	AICc	Log-likelihood	Likelihood ratio	*P*-value
Mass	Null	747.4	−369.7	6.21	0.0339*
	Alternative	743.2	−366.6		
Wing length	Null	1065.1	−528.6	0.205	0.744
	Alternative	1067.0	−474.5		
Baseline CORT	Null	962.0	−477.0	5.02	0.051^{+}
	Alternative	958.9	−474.5		
Stress CORT	Null	1606.0	−798.9	2.26	0.212
	Alternative	1605.7	−797.8		
Dex CORT	Null	1433.6	−712.8	151.4	0.0001***
	Alternative	1284.3	−637.2		
OXY	Null	2024.8	−1008.4	0.37	0.724
	Alternative	2026.4	−1008.2		
d-ROMs	Null	1408.2	−700.1	11.77	0.0024**
	Alternative	1398.5	−694.2		

As such, a significant LRT indicates a trait for which variability differs among populations. The reported *P*-values are adjusted for the false discovery rates ($^{+}0.05 < P < 0.1$; $*P < 0.05$; $**P < 0.01$; $***P < 0.001$).

better fit (LRT, $P < 0.05$ after controlling the false discovery rate), then a given trait exhibited different variability across populations. For morphology, we found that New York and Alaska are similarly variable in terms of wing length ($X^2 = 0.205$, $P = 0.744$), whereas mass was slightly more variable in New York ($X^2 = 6.21$, $P = 0.034$). Post-dex CORT variability appears to be much higher in Alaska ($X^2 = 151.4$, $P < 0.001$). By contrast, the New York population has marginally higher baseline CORT variability ($X^2 = 5.01$, $P = 0.051$), and there was no difference in stress-induced CORT variability ($X^2 = 2.26$, $P = 0.212$). Finally, our two non-endocrine physiological variables exhibited unique variability patterns across populations, with d-ROM variability higher in the New York population ($X^2 = 11.77$, $P = 0.002$) and no difference for OXY ($X^2 = 0.37$, $P = 0.724$).

How does steroid level variability compare to variability in other traits within populations? Using a similar approach as above, we examined whether baseline, stress-induced, and post-dex CORT were similarly variable to other traits (Table 2). We found significantly better fits for the unequal-variance models when comparing baseline CORT to mass (Fig. 1A; $X^2 = 7.4$, $P = 0.016$), OXY (Fig. 1A; $X^2 = 15.2$, $P = 0.001$), and nest provisioning rate (Fig. 1A; $X^2 = 10.8$, $P = 0.006$). In each case, baseline CORT was the more variable trait. However, baseline CORT was statistically no more variable than wing length (Table 2; $X^2 = 0.09$, $P = 0.769$) or d-ROMs (Table 2; $X^2 = 0.2$, $P = 0.767$). Stress-induced CORT was only more variable than antioxidant capacity (Fig. 1B; $X^2 = 7.83$, $P = 0.0007$) and nest provisioning rate (Fig. 1B; $X^2 = 0.79$, $P = 0.438$). As such, models with shared variance best explained the data when comparing stress-induced CORT with mass (Fig. 1B; $X^2 = 2.2$, $P = 0.183$), wing length (Fig. 1B; $X^2 = 1.1$, $P = 0.361$), and d-ROMs (Fig. 1B; $X^2 = 2.3$, $P = 0.178$). Finally, post-dex CORT was more variable than any other trait (Fig. 1C; see Table 2), including other measures of CORT (Table 3) taken at baseline (Fig. 1D; $X^2 = 55.0$, $P < 0.0001$) and following exposure to a standardized stressor (Fig. 1D; $X^2 = 15.2$, $P < 0.0001$).

Endocrine variability on a macroevolutionary scale

We next examined patterns of variability across different traits through a macroevolutionary lens, using data collected from 17 species of *Anolis* lizard (Fig. 2). First, in a series of phylogenetic paired *t*-tests, we found that the normalized CV for T is higher than that of SVL ($t = 5.34$, $P = 0.017$), mass ($t = 3.03$, $P = 0.040$), and locomotion rates ($t = 3.23$, $P = 0.040$). However, the CV for T is indistinguishable from push-up display rate ($t = 1.02$, $P = 0.660$). Mass and locomotion rate were the next-most variable traits, exhibiting similar variability to one another ($t = 0.170$, $P = 0.866$) and greater variability than SVL (mass: $t = 3.20$, $P = 0.040$; locomotion rate: $t = 4.02$, $P = 0.016$; Fig. 2).

Table 2 Fit comparison to test whether a model with equal or unequal variances is a better fit when comparing the distribution of two traits to one another and controlling for effects of population, sex, and measurement error

Comparison	Model	df	AIC	logLik	X^2	*P*-value
CORT (BL)-Mass	Null	5	3564.0	−1776.4	7.4	0.016*
	Alternative	6	3559.4	−1772.7		
CORT (BL)-Wing length	Null	5	3440.1	−1714.1	0.09	0.769
	Alternative	6	34,423.1	−1714.0		
CORT (BL)-OXY	Null	5	3289.7	−1638.8	15.2	0.001**
	Alternative	6	3276.5	−1631.3		
CORT (BL)-dROMs	Null	5	3105.1	−1546.5	0.15	0.767
	Alternative	6	3106.9	−1546.5		
CORT (BL)-Provisioning	Null	4	1895.3	−943.7	10.80	0.006**
	Alternative	5	1886.5	−938.2		
CORT (S)-Mass	Null	5	3415.2	−1701.6	2.18	0.219
	Alternative	6	3414.9	−1700.5		
CORT (S)-Wing	Null	4	3301.4	−1644.7	1.11	0.402
	Alternative	5	3302.2	−1644.1		
CORT (S)-OXY	Null	5	3139.4	−1563.7	7.83	0.016*
	Alternative	6	3133.5	−1559.8		
CORT (S)-dROMs	Null	5	2922.2	−1454.1	2.33	0.219
	Alternative	6	2922.1	−1454.1		
CORT (S)-Provisioning	Null	4	1751.5	−871.7	7.25	0.016*
	Alternative	5	1746.3	−868.1		
CORT (Dex)-Mass	Null	5	3264.4	−1627	103.5	<0.001***
	Alternative	6	3162.9	−1575.4		
CORT (Dex)-Wing	Null	5	3127.3	−1558.7	54.1	<0.0001***
	Alternative	6	3075.2	−1531.6		
CORT (Dex)-OXY	Null	5	3109.2	−1579.6	113.7	<0.0001***
	Alternative	6	2997.5	−1492.7		
CORT (Dex)-dROMs	Null	5	2701.4	−1345.7	41.3	<0.0001***
	Alternative	6	2662.1	−1325		
CORT (Dex)-Provisioning	Null	4	947.6	−469.8	19.4	<0.0001***
	Alternative	5	930.2	−460.1		

The null model entails the default of equal variances among the two traits, while the alternative allows them to remain variable. We infer differences in variability by testing the hypothesis that unequal-variance models better explain the data. FDR-adjusted *P*-values are reported ($*P < 0.05$; $**P < 0.01$; $***P < 0.001$).

With such a comparative dataset we could also examine whether variability evolves in a correlated fashion among traits (Supplementary Table S2). However, normalized CVs were uncorrelated across the phylogeny for every trait pair (summarized in Supplementary Table S1), except for SVL and mass ($F_{1,12} = 41.63$, $P = 0.0002$; Fig. 3).

Discussion

Here, we compared the degree to which steroid hormone levels and other traits vary among individuals across populations and species. Most work describes phenotypes through the perspective of central tendency alone, but we found that steroid hormones exhibit distinct patterns of variation when compared with other traits. We also find that circulating steroid levels, along with measures of oxidative stress, are differentially variable across two populations spanning different biogeographic contexts (temperate vs. subarctic). On the scale of populations, we find that baseline and stress-induced CORT are similarly variable to other traits, while post-dex CORT is significantly more variable than all other phenotypes we

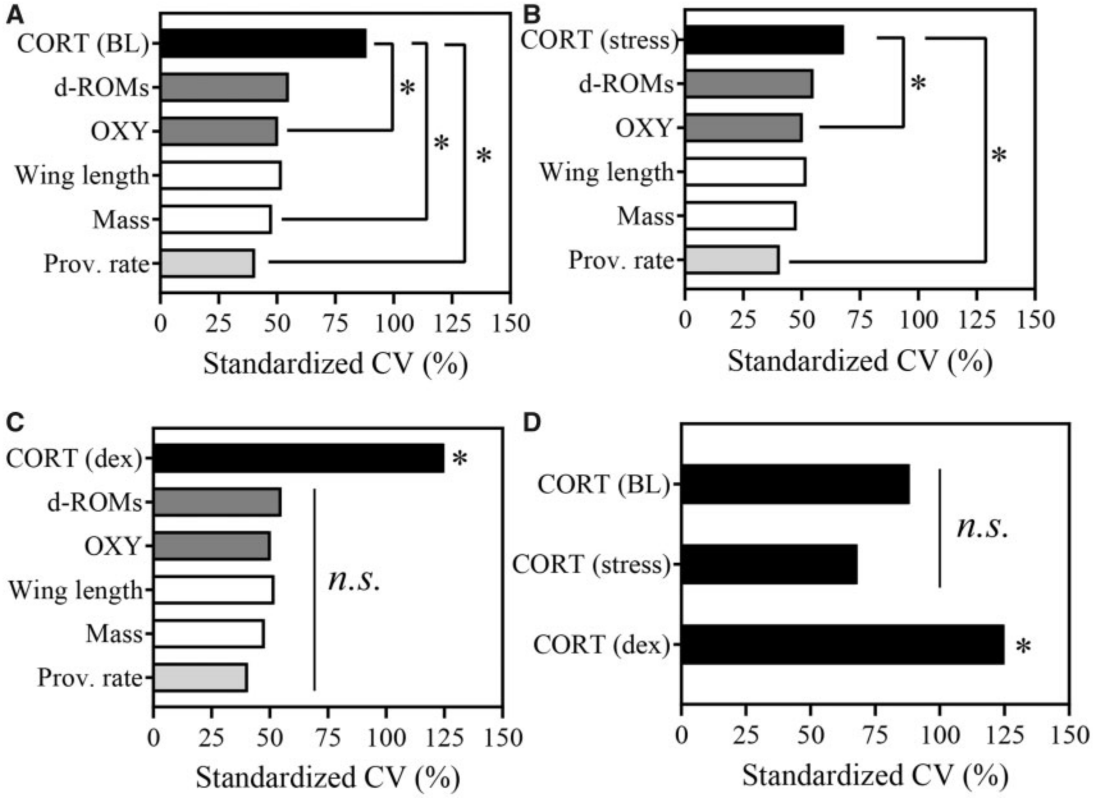

Fig. 1 Pairwise comparisons of within-population variability between baseline CORT (**A**), stress-induced CORT (**B**), and post-dex CORT (**C**) and other traits, plus the three CORT measures compared with one another (**D**). The plotted value is the coefficient of variation (CV) computed after each variable was rescaled to fit minimum = 0 and maximum = 100. Statistically significant comparisons ($P < 0.05$) are denoted with a bracket and asterisk (*) or a single asterisk above the different trait while statistically similar variables are denoted *n.s.* All models accounted for within-individual CV (measurement error or within-individual variation), though it is nonetheless uncorrelated with population CV (B; $R^2 = 0.10$, $P = 0.487$). Note that each panel shows combinations of the same data, but we separated each set of model comparisons into its own panel for clarity.

Table 3 Fit comparison summary for models comparing variability in baseline, stress-induced, and post-dex CORT

Comparison	Model	Df	AIC	logLik	X^2	*P*-value
Baseline–stress	Null	5	3268.3	−1628.1	0.78	0.375
	Alternative	6	3269.5	−1627.8		
Baseline–Dex	Null	5	3089.5	−1539.8	55.0	<0.0001***
	Alternative	6	3036.5	−1512.2		
Stress–Dex	Null	5	2963.0	−1476.5	15.2	<0.0001***
	Alternative	6	2897.6	−1442.8		

The more complex model in each pair allows traits to have unequal variances, so a significant LRT indicates that two traits are differentially variable. *P*-values reported are adjusted to control for FDR.

examined. This suggests that although circulating steroid levels are not extraordinarily variable organismal phenotypes, the sum of the endocrine response may be. In turn, this is consistent with the idea that endocrine systems evolve via modification of multiple components of the system rather than shifting a single element alone (Hau 2007). Meanwhile, we also show that the sex steroid T is as variable as social behavior production across a radiation of tropical lizards. Altogether, these findings show that steroid hormones—at least CORT and T—show levels of within-population variability that are like many other traits known to evolve in response to multiple processes, including selection.

Population-level analyses

We find that within-population variability of many traits differs between tree swallows breeding in New York and Alaska. CORT levels post-dexamethasone treatment were more variable in Alaska, whereas body mass, d-ROMs, and possibly baseline CORT were more variable in New York. Other traits we measured, such as wing length, antioxidant capacity, and stress-induced CORT, showed similar variability in both populations. Overall, these findings suggest that within-population variability in CORT levels is not inherently higher than for other traits; thus, steroid hormone variability is not unilaterally more variable than other traits.

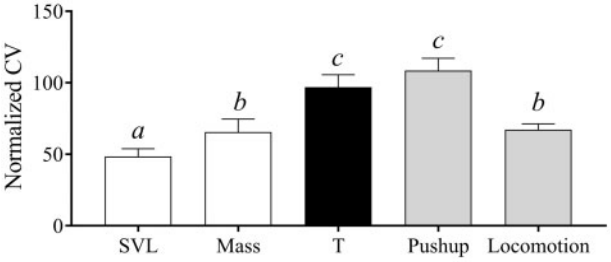

Fig. 2 CV for morphological, endocrine, and behavioral traits in *Anolis* lizards. White bars = morphology (snout–vent length = SVL; body mass = mass); black bar = plasma testosterone (T); gray bars = behavior (push-up courtship display rate = pushup; locomotory activity = locomotion). Note that each bar represents an average taken across species. Normalized CV is computed as CV = std/mean, but only after transforming each distribution to have minimum = 0 and maximum = 100. Letters above each bar indicate statistically similar groups, so that any pair of traits marked with different letters are significantly different ($P < 0.05$) in terms of trait variability.

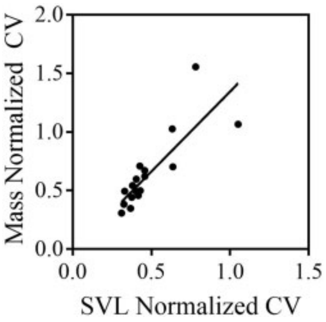

Fig. 3 Among 17 *Anolis* lizard species (each point = 1 species), the only traits exhibiting correlated variability are SVL and mass ($F_{1,12} = 41.6$, $P = 0.0002$). The best-fit line generated from PGLS is shown.

Among the traits whose variability did differ between populations, the most pronounced differences were seen in post-dexamethasone CORT levels, which were higher in Alaska than in New York. In tree swallows, stress resilience—defined as the ability to continue reproducing in the presence of stressors—is greatest in individuals that show both a robust CORT stress response, and strong negative feedback (lower post-dex CORT; Zimmer et al. in review). Because birds breeding in Alaska face both a shorter breeding season—with a reduced probability of re-nesting in the event of reproductive failure—and a less predictable environment, having strong negative feedback may be particularly important for reproductive success in this population. But while having strong negative feedback promotes reproductive success, it could also impair survival (as reproducing under challenging conditions may induce survival costs). Thus, it is possible that selection favors alternative phenotypes in the population, resulting in more within-population variation in this trait than in the relatively relaxed environment in New York. Alternatively, it is possible that post-dex CORT is also more variable *within* individuals than the other traits measured; in this case, the greater environmental variation in Alaska could lead to greater observed within-population variation in this trait, even in the absence of greater inter-individual variation.

Within each tree swallow population, average levels of within-population variability in baseline and stress-induced CORT were higher than variability in some other traits, including body mass, antioxidant capacity, and nest provisioning rate. In contrast to the patterns seen in baseline and stress-induced CORT, post-dex CORT levels were significantly more variable than all other traits measured. If the efficacy of negative feedback (measured by post-dex CORT) is a particularly important component of stress resilience (Taff et al. 2018; Zimmer et al. in review), then we might predict that selection would lead to canalization on this trait. Conversely, as described above, selection could favor alternative phenotypes in the population.

Although behavior is often considered to be a highly-variable animal trait (Scheiner 1993; Gittleman et al. 1996; Wimberger and de Queiroz 1996; Duckworth 2009), we found that variability was higher for all three measures of CORT than for nest provisioning behavior. We suspect that nest provisioning itself exhibits reduced within-population variation compared with other behaviors. This may also be because nest provisioning rates are likely to evolve under stronger stabilizing selection, where variability above the mean is limited by food processing rates and below the mean by offspring mortality. Of course, other behavioral traits may differ in this regard, showing especially high or low levels of individual flexibility (Bell et al. 2009; Duckworth 2009), and thus may compare differently to within-population variability in CORT.

It is important to recognize that these analyses compare variability of traits for which measurements were taken on different time scales. CORT measurements, for instance, are collected at single time points, and circulating CORT levels can change rapidly. On the other hand, provisioning behavior estimates were obtained from daily averages, which reflects an integration that occurs over a much longer period, potentially reducing variability estimates. It is also important to note that for the labile traits measured here, within-population variability likely comes from a combination of within-individual and among-individual variation (Westneat et al. 2015; Hau et al. 2016). Future analyses that separate within-population variability into its component parts, and compare variability metrics across different time scales, could provide important new insights.

Species-level analyses

On a macroevolutionary scale, circulating T levels are among the most variable traits we measured in *Anolis* lizards. The only other trait with similarly high variability across 17 species was pushup display rate, which measures how often an individual produces a courtship or aggressive display (Jenssen 1977), and is likely highly dependent on the immediate social environment of an individual. Indeed, other traits, such as mass, SVL, and locomotion behavior, were significantly less variable than circulating T (and pushup displays). These findings therefore place variability in T levels at the high end, but certainly not to a degree in which it is dramatically greater than other traits that define animal phenotype.

We also find no correlation between variability in T and any of the other traits we measured. This suggests that endocrine variability can evolve independently from variability in other traits (e.g., body mass or SVL) that are well-known to place constraints on the evolution of other social and sexual behaviors (e.g., Podos 2001; Miles et al. 2018). This speaks to the complex design of animal systems and behavior, whereby effects of constraint and physiology on behavioral output can vary in species-specific ways. In particular, one might have expected to observe a positive relationship between species-level individual variability in T and push up display, given that the former is believed to activate the latter (Johnson et al. 2018). Thus, the lack of any link between these two variables at this scale supports a model in which display behavior may evolve independently of circulating T levels *per se* (Husak and Lovern 2014).

Conclusions

In summary, we show that circulating steroid hormone levels (CORT and T) exhibit within-population variation that is greater than some types of traits, but indistinguishable from others. Our analyses suggest there are multiple evolutionary patterns to understand in endocrine variability, possibly depending on the scale of analysis and species under consideration. For example, in tree swallows, we find that variability in CORT is greater than variability in certain morphological and physiological traits, but statistically indistinguishable from others. However, in *Anolis* lizards, T variability is greater than all morphological traits we examined (we did not examine any physiological traits in these species as comparable data across anole species were unavailable). These differences themselves point to something interesting about the nature of hormone system variability, suggesting that this trait could be influenced by one of many factors including evolutionary history (i.e., bird vs. lizard), the hormone in question (CORT vs. T), and the context in which hormones are measured (baseline vs. stress-induced vs. post-dex CORT). The ultimate and proximate causes of such differences in evolutionary patterning of hormone levels merit further attention, particularly since we know little about the macroevolution of different steroid systems. Indeed, understanding this patterning and how it relates to organismal functioning will be vital to further unlocking the mechanisms that underlie phenotypic diversity.

Acknowledgments

We are grateful to the many people who helped to collect these data, particularly David Scheck, Jocelyn Stedman, Dan Ardia, and Alexandra Rose. We thank the Company of Biologists, the SICB DCE, DAB, DCPB, and DEE divisions who sponsored our participation in the SICB 2018 meeting.

Funding

This work was supported by the National Science Foundation (IOS-1655730 to M.J.F., IOS-1457251 to M.N.V., IOS-1257773 and IOS-1656618 to L.B.M., and IOS-1257021 to M.A.J.) and the National Geographic Society (Research and Exploration Grant to J.F.H).

Supplementary data

Supplementary data available at *ICB* online.

References

Adkins-Regan E. 2008. Do hormonal control systems produce evolutionary inertia? Philos Trans R Soc Lond B Biol Sci 363:1599–609.

Baird TA, Hews DK. 2007. Plasma levels of steroid hormones in territorial and non-territorial male collard lizards. Physiol Behav 92:755–63.

Bell AM, Hankison SJ, Laskowski KL. 2009. The repeatability of behaviour: a meta-analysis. Anim Behav 77:771–83.

Bennett AF. 1987. Interindividual variability: an underutilized resource. In: Feder ME, Burggren WW, Bennett AF, editors. New directions in ecological physiology. Cambridge: Cambridge University Press. p. 147–69.

Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717–45.

Breuner CW, Delehanty B, Boonstra R. 2013. Evaluating stress in natural populations of vertebrates: total CORT is not good enough. Funct Ecol 27:24–36.

Bridge ES, Bonter DN. 2011. A low-cost radio frequency identification device for ornithological research: low-cost RFID reader. J Field Ornithol 82:52–9.

Cockrem JF. 2013. Corticosterone responses and personality in birds: individual variation and the ability to cope with environmental changes due to climate change. Gen Comp Endocrinol 190:156–63.

Costantini D, Dell'Omo G. 2006. Environmental and genetic components of oxidative stress in wild kestrel nestlings (*Falco tinnunculus*). J Comp Physiol B Biochem Syst Environ Physiol 176:575–9.

Crews D, Moore MC. 1986. Evolution of mechanisms controlling mating behavior. Science 231:121–5.

Duckworth R. 2009. The role of behavior in evolution: a search for mechanism. Evol Ecol 23:513–31.

Dunning J. 1992. CRC handbook of avian body masses. Boca Raton (FL): CRC Press.

Freeman S, Herron JC. 2007. Evolutionary analysis. New Jersey: Pearson Prentice Hall.

Fuxjager MJ, Miles MC, Goller F, Petersen J, Yancey Y. 2017. Androgens support male acrobatic courtship behavior by enhancing muscle speed and easing the severity of its tradeoff with force. Endocrinology 158:4038–688.

Fuxjager MJ, Miles MC, Schlinger BA. 2018. Evolution of the androgen-induced male phenotype. J Comp Physiol A 204:81–70.

Gittleman JL, Anderson CG, Kot M, Luh HK. 1996. Phylogenetic lability and rates of evolution: a comparison of behavioral, morphological and life history traits. In: Martins EP, editor. Phylogenies and the comparative method in animal behavior. New York (NY): Oxford University Press. p. 166–205.

Gleason ED, Fuxjager MJ, Oyegbile TO, Marler CA. 2009. Testosterone release and social context: when it occurs and why. Front Neuroendocrinol 30:460–9.

Grant P. 1999. Ecology and evolution of Darwin's finches. New Jersey: Princeton University Press.

Hau M. 2007. Regulation of male traits by testosteronse: implications for the evolution of vertebrate life histories. Bioessays 29:133–44.

Hau M, Casagrande S, Ouyang JQ, Baugh AT. 2016. Glucocorticoid-mediated phenotypes in vertebrates: multilevel variation and evolution. In: Naguib M, Mitani JC, Simmons LW, Barrett L, Healy S, Zuk M, editors. Advances in the study of behavior. Amsterdam: Elsevier Academic Press. p. 41–115.

Hau M, Ricklefs RE, Wikelski M, Lee KA, Brawn JD. 2010. Corticosterone, testosterone and life-history strategies of birds. Proc R Soc B Biol Sci 277:3203–12.

Hirschenhauser K, Oliveira RF. 2006. Social modulation of androgens in male vertebrates: meta-analyses of the challenge hypothesis. Anim Behav 71:265–77.

Holmes MM, Bartrem CL, Wade J. 2007. Androgen dependent seasonal changes in muscle fiber type in the dewlap neuromuscular system of green anoles. Physiol Behav 91:601–8.

Holmes MM, Wade J. 2005. Testosterone regulates androgen receptor immunoreactivity in the copulatory, but not courtship, neuromuscular system in adult male green anoles. J Neuroendocrinol 17:560–9.

Husak JF, Lovern MB. 2014. Variation in steroid hormone levels among Caribbean *Anolis* lizards: endocrine system convergence? Horm Behav 65:408–15.

Jenkins BR, Vitousek MN, Hubbard JK, Safran RJ. 2014. An experimental analysis of the heritability of variation in glucocorticoid concentrations in a wild avian population. Proc R Soc B Biol Sci 281:20141302.

Jenssen TA. 1977. Evolution of anoline lizard display behavior. Am Zool 17:203–15.

Johnson MA, Caton JL, Cohen RE, Vandecar JR, Wade J. 2010. The burden of motherhood: the effect of reproductive load on female lizard locomotor, foraging, and social behavior. Ethology 116:1217–25.

Johnson MA, Kircher BK, Castro DJ. 2018. The evolution of androgen receptor expression and behavior in *Anolis* lizard forelimb muscles. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 204:71–9.

Johnson MA, Wade J. 2010. Behavioural display systems across nine *Anolis* lizard species: sexual dimorphisms in structure and function. Proc R Soc B Biol Sci 277:1711–9.

Kempenaers B, Peters A, Foerster K. 2008. Sources of individual variation in plasma testosterone levels. Philos Trans R Soc Lond Ser B Biol Sci 363:1711–23.

Ketterson ED, Atwell JW, McGlothlin JW. 2009. Phenotypic integration and independence: hormones, performance, and response to environmental change. Integr Comp Biol 49:365–79.

King RB, Cline JH, Hubbard CJ. 2004. Heritable variation in testosterone levels in male garter snakes (*Thamnophis sirtalis*). J Zool 264:143–7.

Langkilde T, Shine R. 2006. How much stress do researchers inflict on their study animals? A case study using a scincid lizard, *Eulamprus heatwolei*. J Exp Biol 209:1035–43.

Losdat S, Helfenstein F, Blount JD, Richner H. 2014. Resistance to oxidative stress shows low heritability and high common environmental variance in a wild bird. J Evol Biol 27:1990–2000.

MacColl ADC, Hatchwell BJ. 2003. Heritability of parental effort in a passerine bird. Evolution 57:2191–5.

McGlothlin JW, Whittaker DJ, Schrock SE, Gerlach NM, Jawor JM, Snajdr EA, Ketterson ED. 2010. Natural selection on testosterone production in a wild songbird population. Am Nat 175:687–701.

Miles MC, Schuppe ER, Ligon RM, Fuxjager MJ. 2018. Macroevolutionary patterning of woodpecker drums reveals how sexual selection elaborates signals under constraint. Proc R Soc B 285:20172628.

Olsson M, Wilson M, Isaksson C, Uller T, Mott B. 2008. Carotenoid intake does not mediate a relationship between reactive oxygen species and bright colouration: experimental test in a lizard. J Exp Biol 211:1257–61.

Podos J. 2001. Correlated evolution of morphology and vocal signal structure in Darwin's finches. Nature 409:185–8.

Revell L. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3:217–23.

Sapolsky RM, Romero LM, Munck AU. 2000. How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. Endocr Rev 21:55–89.

Scheiner SM. 1993. Genetics and evolution of phenotypic plasticity. Annu Rev Ecol Syst 24:35–68.

Stedman JM, Hallinger KK, Winkler DW, Vitousek MN. 2017. Heritable variation in circulating glucocorticoids and endocrine flexibility in a free-living songbird. J Evol Biol 30:1724–35.

Taff CC, Zimmer C, Vitousek MN. 2018. Efficacy of negative feedback in the HPA axis predicts recovery from acute challenges. Biol Lett 14 (doi: 10.1098/rsbl.2018.0131).

Taff CC, Vitousek MN. 2016. Endocrine flexibility: optimizing phenotypes in a dynamic world? Trends Ecol Evol 31:476–88.

Tsuji JS, Huey RB, van Berkum FH, Garland T, Shaw RG. 1989. Locomotor performance of hatchling fence lizards (*Sceloporus occidentalis*): quantitative genetics and morphometric correlates. Evol Ecol 3:240–52.

Vitousek MN, Taff CC, Ardia DR, Stedman JM, Zimmer C, Salzman TC, Winkler DW. In revision. The lingering impact of stress: brief acute glucocorticoid exposure has sustained, dose-dependent effects on reproduction. Proc R Soc B Biol Sci (https://doi.org/10.5061/dryad.s24121t).

Vitousek MN, Taff CC, Hallinger KK, Zimmer C, Winkler DW. 2018. Hormones and fitness: evidence for trade-offs in glucocorticoid regulation across contexts. Front Ecol Evol 6:1–14.

Vitousek MN, Tomasek O, Albrecht T, Wilkins MR, Safran RJ. 2016. Signal traits and oxidative stress: a comparative study across populations with divergent signals. Front Ecol Evol 4:56.

Westneat DF, Wright J, Dingemanse NJ. 2015. The biology hidden inside residual within-individual phenotypic variation. Biol Rev 90:729–43.

Williams TD. 2008. Individual variation in endocrine systems: moving beyond the ‘tyranny of the Golden Mean’. Philos Trans R Soc Lond B Biol Sci 363:1687–98.

Wimberger PH, de Queiroz A. 1996. Comparing behavioral and morphological characters as indicators of phylogeny. In Martins EP, editor. Phylogenies and the comparative method in animal behavior. New York (NY): Oxford University Press. p. 206–33.

Wingfield JC, Hegner RE, Dufty AM, Ball GF. 1990. The “challenge hypothesis”: theoretical implications for patterns of testosterone secretion, mating systems, and breeding strategies. Am Nat 136:829–46.

Wingfield JC, Jacobs JD, Tramontin AD, Perfito N, Meddle S, Maney DL, Soma KK. 2000. Toward an ecological basis of hormone–behavior interactions in reproduction of birds. In: Wallen K, Schneider JE, editors. Reproduction in context. Cambridge: MIT Press. p. 85–128.

Wingfield JC, Sapolsky RM. 2003. Reproduction and resistance to stress: when and how. J Neuroendocrinol 15:711–24.

Wingfield JC, Visser ME, Williams TD. 2008. Integration of ecology and endocrinology in avian reproduction: a new synthesis. Philos Trans R Soc Lond B Biol Sci 363:1581–8.

Zimmer C, Taff CC, Ardia DR, Winkler DW, Vitousek MN. In review. On again, off again: acute stress response and negative feedback together predict resilience to experimental stressors.