Review

Individual variation in endocrine systems: moving beyond the ‘tyranny of the Golden Mean’

Tony D. Williams*

Department of Biological Sciences, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6

Twenty years ago, Albert Bennett published a paper in the influential book *New directions in ecological physiology* arguing that individual variation was an ‘underutilized resource’. In this paper, I review our state of knowledge of the magnitude, mechanisms and functional significance of phenotypic variation, plasticity and flexibility in endocrine systems, and argue for a renewed focus on inter-individual variability. This will provide challenges to conventional wisdom in endocrinology itself, e.g. re-evaluation of relatively simple, but unresolved questions such as structure–function relationships among hormones, binding globulins and receptors, and the functional significance of absolute versus relative hormone titres. However, there are also abundant opportunities for endocrinologists to contribute solid mechanistic understanding to key questions in evolutionary biology, e.g. how endocrine regulation is involved in evolution of complex suites of traits, or how hormone pleiotropy regulates trade-offs among life-history traits. This will require endocrinologists to embrace the raw material of adaptation (heritable, individual variation and phenotypic plasticity) and to take advantage of conceptual approaches widely used in evolutionary biology (selection studies, reaction norms, concepts of evolutionary design) as well as a more explicit focus on the endocrine basis of life-history traits that are of primary interest to evolutionary biologists (cf. behavioural endocrinology).

Keywords: endocrine systems; inter-individual variation; reaction norms; heritability; plasticity

It is difficult to understand why statisticians [and endocrinologists/physiologists] commonly limit their inquiries to Averages, and do not revel in more comprehensive views. Their souls seem as dull to the charm of variety as that of the native of one of our flat English counties, whose retrospect of Switzerland was that, if its mountains could be thrown into its lakes, two nuisances would be got rid of at once

(Galton 1889)

1. INTRODUCTION

Twenty years ago, Bennett (1987) published a paper in the influential book *New directions in ecological physiology* (Feder et al. 1987) highlighting the almost complete focus on central tendency in physiological studies, which he described as the ‘tyranny of the Golden Mean’. Even 20 years ago, this was hardly a new message as evidenced by the opening quote in this paper made nearly 100 years earlier by Sir Francis Galton, F.R.S., in his book *Natural inheritance* (Galton 1889). Bennett (1987) suggested that analysis of inter-individual variability could help bridge physiological studies to other fields of biology such as ecology, behaviour, evolution and genetics (see also Feder et al. 2000). Feder et al.’s (1987) book led to significant advances in several ‘new directions’ in ecological physiology, e.g. selection studies (Swallow & Garland 2005 and references therein) and phylogenetically based approaches (Garland et al. 2005). Furthermore, many other biological disciplines have embraced the analysis of individual variation in areas such as behavioural ecology, population biology, quantitative genetics and epidemiology (e.g. Bolnick et al. 2003; Lloyd-Smith et al. 2005; Breckling et al. 2006; Reale et al. 2007). Given the importance of inter-individual variation to the integration of endocrinology and evolutionary biology, how have (avian) comparative and ecological endocrinologists risen to Bennett’s (1987) challenge in the last 20 years? To what extent do we understand, and can we explain, the magnitude, patterns, causes and consequences of inter-individual variation in endocrine-regulatory networks and endocrine-mediated traits? It is increasingly widely recognized that, perhaps more so than any other physiological system, hormones are critically involved in adaptation and evolution of complex traits (Ketterson & Nolan 1999; Sinervo 1999; Zera & Harshman 2001; Sih et al. 2004a). Evolutionary changes in endocrine regulation are thought to be an especially important mechanism by which entire suites of traits evolve in a coordinated manner in response to environmental change, either via selection on heritable, fitness-related, individual variation or through adaptive phenotypic plasticity (Dufty et al. 2002; Zera et al. 2007). Endocrine systems, and hormone titres in particular, display marked, but poorly understood, inter-individual
variability, phenotypic plasticity and (reversible) phenotypic flexibility (*sensu* Piersma & Drent 2003). Therefore, comparative and ecological endocrinologists are in a unique position to capitalize on the central role that hormones play in adaptation, and to make significant contribution to both evolutionary biology and our understanding of the basic functioning of endocrine systems themselves—provided that we are prepared to view endocrine systems from novel perspectives and to take advantage of novel analytical and conceptual approaches.

2. HOW LARGE IS INTER-INDIVIDUAL VARIATION IN ENDOCRINE TRAITS?

Although it is a widely held view that there is plenty of physiological diversity among individuals of the same species (e.g. Adkins-Regan 2005, p. 179), it is still very rare for authors to present, let alone formally analyse, inter-individual variation (but see Spicer & Gaston 1999). Among a total of 109 figures or tables from 57 research articles in *General and Comparative Endocrinology* (2006), only 9 (8%) presented full inter-individual data as either scatter plots, ranges or 95% CIs. In *Physiological Biochemistry and Zoology*, 168 figures and tables from 40 research papers (2005) included 24 with full inter-individual data (14%); a difference that perhaps reflects the subtitle of this latter journal: *Ecological and Evolutionary Approaches*. The most common form of data presentation was 'mean ± s.e.m., n = 5' confirming that most authors continue to use variance in their data simply to provide confidence limits about the estimated mean values (Bennett 1987), rather than exploring this variance itself. Thus, there is a paucity of studies on inter-individual variation with which to work, but what can we conclude about the magnitude of inter-individual variation in hormone titres and endocrine-mediated traits based on the relatively small number of studies that are available?

Hormones titres are inherently variable, showing marked diurnal, age-dependent and life-history variation (e.g. Finch & Rose 1995; Norris 1997), and variation due to changes in physiological state (e.g. breeding versus non-breeding, stressed versus non-stressed). Clearly, absolute standardization will be difficult, especially in field studies, but this should not be used as another reason to ignore variability (although the issue of 'natural' episodic hormone release in generating apparent variability in single blood samples clearly needs to be better addressed). Indeed, a finding of repeatability (see below) against a background of random events that cannot be controlled by the researcher would be of particular interest. Thus, if measurement conditions are standardized as far as possible, we can get at least a rough estimate of the extent of inter-individual variation for a given developmental, ontogenetic, seasonal or physiological state. The few data that are available from carefully selected studies (table 1) confirm that there is very large variation in endocrine traits with hormone titres varying 5- to 15-fold among individuals for a given physiological state in both free-living and captive animals (see also Kempenaers et al. 2008). Variation in hormone titres appears to be greater than for many other physiological traits, with the possible exception of other humoral components (e.g. triglycerides, carotenoids; table 1). There can also be marked individual variation in time-dependent patterns of endocrine traits, e.g. in rainbow trout (*Oncorhynchus mykiss*), individual variation in peaks in thyroid hormone levels were irregular and asynchronous without apparent influence of day/night, sex or feeding (Gomez et al. 1997). Time-dependent changes in plasma corticosterone levels in response to a standardized stressor, an indicator of hypothalamo-pituitary-adrenal (HPA) axis reactivity, are also highly variable (Cockrem & Silverin 2002). In mammals, Guimont & Wynne-Edwards (2006) reported that 'remarkably few individuals showed the average' stress response of a 50 ng ml⁻¹ cortisol increase (table 1) followed by a recovery to baseline; individual variation was independent of sex, age, body mass or housing conditions; and hierarchical cluster analysis could not partition between-individual variation confirming a continuum of variation. At longer time scales, inter-individual variation in age-dependent changes in hormone levels is also evident, e.g. in human females, Ferell et al. (2005) showed that population-level analyses of steroid hormone and menstrual cycle changes do not accurately predict relationships within and among individual women. Thus, the principles of homeostasis notwithstanding, large (5- to 10-fold) inter-individual variation is the norm for hormone titres and some endocrine-mediated traits at all temporal scales. What are the analytical and biological implications of a full recognition of this variation?

3. IS INTER-INDIVIDUAL VARIATION BIOLOGICAL? MEASUREMENT ERROR AND REPEATABILITY

One of the main reasons that inter-individual variation has been 'ignored' is the concern that observed variability is due to measurement error (Bennett 1987) resulting from the stochastic nature of endocrine assays and 'procedural' errors (e.g. spillage, evaporation of samples). However, measurement error is typically less than 5-10% (rarely 15%) for most plasma hormone and plasma metabolite assays and this cannot explain the magnitude of variability reported for hormone titres and other humoral components in table 1. A related concern is that extreme trait values are atypical or abnormal (Bennett 1987) or are artefacts of physiological manipulations or measurement. However, as figure 1 illustrates, this is really a subjective assessment where researchers need, at least, to bear the burden of proof. In fact, individuals with such 'extreme phenotypes' could be very informative in understanding links between mechanism and phenotypic variation. For example, in figure 1, in the context of functional hypotheses relating hematocrit and oxygen-carrying capacity to metabolically demanding reproductive effort, how does the 'extreme phenotype' individual function with such low hematocrit?

Repeatability measures the extent to which an individual's phenotypic trait value remains consistent over time (Bennett 1987; Dohm 2002). When correctly estimated (Lessells & Boag 1987), repeatability...
estimates can provide insight into the heritable nature of traits and their potential response to selection (Dohm 2002; see below). Despite the value of this estimate, our knowledge of repeatability of physiological traits in general, and especially endocrine traits, remains surprisingly poor (cf. Dohm 2002). A few studies have reported

Table 1. Some examples of the magnitude of inter-individual variation for plasma hormone titres, other humoral components and other physiological traits.

<table>
<thead>
<tr>
<th>physiological trait</th>
<th>range</th>
<th>difference</th>
<th>standardized conditions or state and species</th>
<th>captive/ free living</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>hormone titres</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17β-oestradiol</td>
<td>0.2–2.2 ng ml⁻¹</td>
<td>11-fold</td>
<td>egg production, zebra finch</td>
<td>C</td>
<td>E. Wagner (2007; personal communication)</td>
</tr>
<tr>
<td>Testosterone</td>
<td>44–423 pg ml⁻¹</td>
<td>10-fold</td>
<td>follicle development, starling</td>
<td>F</td>
<td>Williams et al. (2004)</td>
</tr>
<tr>
<td></td>
<td>98–383 pg ml⁻¹</td>
<td>4-fold</td>
<td>egg production, starling</td>
<td>F</td>
<td>Williams et al. (2004)</td>
</tr>
<tr>
<td></td>
<td>1.8–11.9 ng ml⁻¹</td>
<td>6-fold</td>
<td>early breeding, male junco</td>
<td>F</td>
<td>Jawor et al. (2006)</td>
</tr>
<tr>
<td>Prolactin</td>
<td>3–25 ng ml⁻¹</td>
<td>8-fold</td>
<td>osmotic challenge, tilapia</td>
<td>C</td>
<td>Seale et al. (2006)</td>
</tr>
<tr>
<td>Corticosterone</td>
<td>20–100 ng ml⁻¹</td>
<td>5-fold</td>
<td>standard stressor, trout</td>
<td>C</td>
<td>Schjolden et al. (2005)</td>
</tr>
<tr>
<td></td>
<td>0.6–10.4</td>
<td>15-fold</td>
<td>baseline, non-manipulated great tit</td>
<td>C</td>
<td>Cockren &amp; Silverin (2002)</td>
</tr>
<tr>
<td></td>
<td>1.8–46.5 pg ml⁻¹</td>
<td>25-fold</td>
<td>baseline, non-manipulated, starling</td>
<td>F</td>
<td>Love et al. (2004)</td>
</tr>
<tr>
<td></td>
<td>10–60 ng ml⁻¹</td>
<td>6-fold</td>
<td>baseline, incubation, eider</td>
<td>F</td>
<td>Bourgeon et al. (2006)</td>
</tr>
<tr>
<td><strong>other humoral components</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Triglycerides</td>
<td>0.3–3.0 mmol l⁻¹</td>
<td>10-fold</td>
<td>mass-corrected</td>
<td>F</td>
<td>T.D. Williams (2004; unpublished data)</td>
</tr>
<tr>
<td>Carotenoids</td>
<td>various a</td>
<td>5- to 6-fold</td>
<td>arrival date, incubation stage</td>
<td>F</td>
<td>Ninni et al. (2004) and Guimont &amp; Wynne-Edwards (2006)</td>
</tr>
<tr>
<td>Heat shock proteins</td>
<td>various a</td>
<td>less than 4-fold</td>
<td>baseline, incubation, eider</td>
<td>F</td>
<td>Guimont &amp; Wynne-Edwards (2006)</td>
</tr>
<tr>
<td>Total anti-oxidants</td>
<td>0.85–2.7 mM</td>
<td>less than 3-fold</td>
<td>incubation stage, blue tits</td>
<td>F</td>
<td>Tummeleht et al. (2006)</td>
</tr>
<tr>
<td><strong>other physiological traits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basal metabolic rate</td>
<td>various a</td>
<td>2-fold</td>
<td>mass-corrected</td>
<td>C</td>
<td>Chappell et al. (1999) and Russell &amp; Chappell (2007)</td>
</tr>
<tr>
<td>Mass-specific DEE</td>
<td>various a</td>
<td>less than 2-fold</td>
<td>chick-provisioning</td>
<td>F</td>
<td>Williams &amp; Vezina (2001)</td>
</tr>
<tr>
<td>Heart rate</td>
<td>300–420 bpm</td>
<td>less than 2-fold</td>
<td>Daphnia</td>
<td>C</td>
<td>Spicer &amp; Gaston (1999)</td>
</tr>
<tr>
<td>Muscle enzyme conc.</td>
<td>various a</td>
<td>3-fold</td>
<td>Hyla</td>
<td>F</td>
<td>James et al. (2005)</td>
</tr>
<tr>
<td>Electrical signal duration</td>
<td>60–135 ms</td>
<td>2-fold</td>
<td>breeding, female, <em>Pollimyrus</em></td>
<td>C</td>
<td>Crawford (1992)</td>
</tr>
</tbody>
</table>

* Includes multiple traits and/or multiple studies.

Figure 1. Outlier or natural extreme phenotype? Most researchers would exclude the individual with the hematocrit value less than 35% in (a) as a statistical ‘outlier’; however, this individual laid two clutches of six eggs (b) with the largest mean egg size.
consistency of individual variation in hormone titres, e.g. stress-induced corticosterone in birds (Cockrem & Silverin 2002) and fishes (Schjolden et al. 2005), and timing of luteinizing hormone (LH) surges, but not peak plasma LH levels in rats (Gans & McClintock 1993; although repeatability was not calculated explicitly in these studies). Several studies have also reported consistency of hormonal responses to a standardized, exogenous hormone treatment, i.e. individual variation in sensitivity to an endocrine stimulus. In male dark-eyed juncos (Junco hyemalis), testosterone (T) release in response to a standard gonadotrophin-releasing hormone (GnRH) challenge was repeatable (r=0.36), and initial (baseline) plasma LH levels predicted post-challenge LH levels (Jawor et al. 2006). Similarly, in non-breeding female zebra finches, inter-individual variation in plasma yolk precursor levels in response to exogenous 17β-oestradiol treatment is consistent among individuals (T.D. Williams 2004, unpublished data).

Other studies have reported repeatability of inter-individual variation in putative endocrine-mediated traits, e.g. oestrogen-dependent yolk precursor production, over time scales of several months (r=0.5–0.7; Challenger et al. 2001).

These studies, though limited in number, and the types of traits that have been investigated suggest that repeatability can vary among traits, and for the same trait in different species. More studies of repeatability and multiple measurements of the same trait within individuals would allow resolution of apparently contradictory findings. Are these systematic differences in repeatability either for the different traits or for the same traits in different species? Has natural selection maintained phenotypic plasticity or flexibility in some physiological systems but not others, and if so why?

Hormone titres are an archetypal example of a phenotypically flexible trait (sensu Piersma & Drent 2003), i.e. they show non-developmental, continuous, but reversible variation within single individuals—yet hormone data are rarely considered in this context. Does selection favour individuals which can more rapidly up- or downregulate hormone titres or individuals which can minimize time lags for these changes (e.g. Sih et al. 2004b)? Are there ‘costs’ generated by the plastic nature of hormone systems per se? One further problem with ignoring repeatability of physiological trait values is that researchers cannot be sure if the physiological measurements (e.g. hormone titres) they obtain truly characterize the phenotype(s) of the sampled individual. Nevertheless, most studies assume this is the case and they then go on to interpret this ‘phylogenetic’ variation functionally and to test adaptive hypotheses, an approach that is increasingly common in evolutionary endocrinology (Zera et al. 2007).

4. ANALYTICAL ISSUES: REPEATED MEASURES DESIGNS AND REACTION NORMS

The simplest analytical method to both deal with and take advantage of inter-individual variation is to use a repeated measures design where multiple measurements are made on the same set of individuals, e.g. during both control or sham and experimental treatments. Each individual acts as its own control and data can be analysed as a change in trait value relative to each individual’s initial value thus controlling for any marked variability in initial values. This straightforward experimental design is still rarely used in endocrine studies and is undoubtedly complicated in certain study systems (e.g. field studies) where any individual is only caught once, or even in laboratory situations where multiple sampling might not easily be approved by institutional animal care committees. However, this does not negate the value and importance of such an approach and it is therefore very important to identify field study systems where repeated measures data can be obtained (e.g. long-term population studies with high natal philopatry or nest-site fidelity) and to provide a scientifically sound rationale for this approach in laboratory studies (the comparison of repeated measures responses for the same species in the field and in captivity is likely to be particularly informative, e.g. Blondel et al. 1999). At the very least, this can then provide a ‘benchmark’ which can be applied to other studies where this approach is less feasible. Repeated measures designs can often reveal significant systematic patterns of variation that can be masked when cross-sectional analyses are used (figure 2). As an example, in studies investigating metabolic adjustments to egg production, Vezina et al. (2006) found no difference in mean daily energy expenditure (DEE) among non-breeding and egg-laying life stages. However, there was marked inter-individual variation in the change in DEE between stages (−33 to +46%), i.e. some individuals markedly increased DEE and others decreased DEE in response to egg production but these changes cancelled out at the population level. Furthermore, this inter-individual variation in change in DEE was systematically related to other traits (e.g. food intake) and, importantly, this variation was repeatable (F. Vezina, unpublished data), i.e. females appear to employ energy management strategies that are highly variable among individuals but consistent within individuals. None of these complexities would have been apparent had we not looked beyond the ‘Golden Mean’ and, more specifically, if we had not obtained repeated measurements on the same set of individuals.

If repeated measurements of individual hormonal or hormone-dependent trait values and the change in trait values are plotted for different environments—rather than simply means and variance—the result is a physiological reaction norm (figure 3). Reaction norms, defined as the set of phenotypes that can be produced by an individual or genotype in different environments, are a central and unifying concept in evolutionary biology (e.g. Schlichting & Pigliucci 1998). However, with one major exception—that of thermal reaction norms for traits such as growth or locomotory behaviour (Angilletta et al. 2003; Kingsolver et al. 2006)—reaction norm approaches have not been widely applied to analysis of endocrine (or other physiological) traits. Clearly, any trait plotted on the y-axis for reaction norms must have a physiological basis, although physiological mechanisms underlying reaction norms remain a major unresolved issue (Angilletta et al. 2003). Thus, a reaction norm approach can be applied to variation in hormone-dependent traits. However, the y-axis trait in a reaction
norm could also be the hormone titre itself, e.g. androgen responsiveness (AR, the ratio of breeding season maximum and baseline androgen titre, e.g. Hirschenhauser et al. 2003) could be treated as a reaction norm if this is calculated for individuals rather than species (see also fig. 1 in Cockrem & Silverin (2002) and fig. 1 in Angelier et al. (2007)). It also seems plausible to plot variation in a hormone-dependent phenotype not against change in ‘external’ environment (e.g. temperature, pH) but in relation to change in ‘internal’ environment, e.g. hormonal state (figure 3b). Here, the reaction norm would indicate plasticity in sensitivity or response of a hormone-dependent trait to changes in endocrine signalling (which itself could be environment dependent). In addition to changes in absolute hormone levels, reaction norms could also capture inter-individual variation in the timing component of hormonal responses: the x-axis could represent time in minutes to hours (as with variation in the stress response, e.g. Cockrem & Silverin (2002)) or developmental time (see Suzuki & Nijhout (2006) for an endocrine example in invertebrates). Alternatively, the y-axis could represent inter-individual variation in circadian timing of some peak circulating-hormone level or the rate of increase in hormone levels (responsivity). Simple, linear reaction norms generate two properties that can be used to characterize and analyse inter-individual variation beyond simple means and variance: the expected trait value in the average environment and the change in trait value with a unit change in environment (fig. 1 in Brommer et al. (2005)). This would allow endocrinologists to explicitly analyse variation in hormonal plasticity in relation to traits that might affect performance and fitness. It is likely that many endocrine and physiological traits will generate more complex, continuous, nonlinear reaction norms, e.g. the change from baseline, pre-breeding hormone levels, to ‘peak’ level breeding levels, and return to lower post-breeding with hormones of the reproductive axis, or the increase then decrease in plasma corticosterone levels due to an acute stressor. Nevertheless, techniques are becoming available for the rigorous analyses of these more complex reaction norms (Izem & Kingsolver 2005; Nussey et al. 2007).

5. IS INTER-INDIVIDUAL VARIATION IN HORMONE TITRES FUNCTIONALLY SIGNIFICANT?
One potential reason why endocrinologists, in particular, have not considered inter-individual variation is the
idea that hormone titles can be functionally uninformative since endocrine regulation occurs mainly through variation in binding globulin action, hormone receptor expression, density or affinity, or intracellular signalling pathways (Norris 1997; Ball & Balthazart 2008); this view is reinforced by increasingly reductionist thinking with a focus on cellular and molecular mechanism. This contrasts with the fact that so much effort in vertebrate endocrinology continues to be directed towards measurement of hormone titles and, interestingly, this also contrasts with invertebrate studies where a predominant focus on hormone titles is the consequence of a large body of evidence implicating regulation of phenotypic trait expression by changes in circulating hormone levels (e.g. growth, polymorphisms; Zera et al. 2007). Yet it remains unclear to what extent receptors or other components of endocrine signalling modulate, contribute to, or override hormone titles in determining hormone-dependent phenotypic trait variation. Selection studies selecting directly (and solely) on circulating hormone levels have demonstrated correlated responses to selection in putative hormone-mediated traits, confirming the functional significance of hormone titles per se. In Japanese quail (Coturnix coturnix japonica), selection for low or high stress-induced corticosterone leads to changes in behavioural phenotype: greater avoidance and more fear-related behaviour in high-selection lines (Jones et al. 1994). Zebra finches (Taeniopygia guttata) selected for high stress-induced corticosterone levels (Evans et al. 2006) showed reduced spatial ability and lower hippocampal miner-alocorticoid-receptor mRNA expression compared with control lines (Hodgson et al. 2007). Other studies have selected on putative endocrine-mediated physiological traits and have shown correlated responses to selection in plasma hormone levels. House mice (Mus domesticus) selectively bred for high levels of voluntary wheel-running behaviour (2.7-fold higher than controls) had baseline plasma corticosterone levels twofold higher than controls, consistent with corticosterone’s function in mobilizing energy during sustained activity. Similarly, in Japanese quail, 32 generations of divergent selection for yolk precursor production, an oestrogen-dependent trait, led to a fourfold decrease and tenfold increase in circulating yolk precursor levels in the low- and high-selection lines, respectively, and this was associated with a corresponding 40% decrease and twofold increase in plasma 17β-oestradiol levels (Chen et al. 1999). Given the integrated nature of endocrine systems, it is likely that selection on hormone titles will also generate correlated responses in other endocrine components (e.g. Hodgson et al. 2007), but this is not always the case. In mice divergently selected for high and low body growth, serum insulin-like growth factor-1 (IGF-I) levels were dramatically increased in all high-selection lines compared with the respective low-selection lines. However, there were no clear patterns of correlated responses to selection for serum IGF-binding proteins or tissue-level gene expression of autocrine/paracrine IGF-I components suggesting an important role of endocrine (i.e. plasma) IGF-I levels in regulation of post-natal growth (Hoeftlisch et al. 2004).

Other studies support the idea that hormone titles per se are functionally important. Individual variation in ‘baseline’ hormone titles can be correlated with, or predict, individual variation in elevated hormone titles, for the same or different hormone, associated with changes in physiological state, as when the endocrine system responds to some environmental stimulus. This suggests that phenotypic variation in the underlying endocrine machinery that determines baseline hormone titles also influences responsiveness of these systems (and perhaps this baseline machinery is the target of selection). In dark-eyed juncos, the response to a GnRH challenge in terms of elevated plasma LH and T correlates with initial baseline titles for both these hormones, i.e. despite marked seasonal changes in GnRH responsiveness, individual variation in short-term increases in LH or T remain consistent over time (Jawor et al. 2006). Similarly, in hamsters post-stress, corticosterone levels are positively correlated with pre-stress baseline corticosterone levels (Guimont & Wynne-Edwards 2006). These studies suggest that individual variation in baseline, non-manipulated hormone levels can be informative in terms of individual variation in subsequent activation of endocrine axes due to hypothalamic signalling or stress, respectively. Thus, these studies strongly support the idea that inter-individual variation in circulating plasma hormone levels is functionally significant and, more importantly, that when hormone titles themselves are the target of selection this can lead to correlated changes in downstream components of endocrine regulation such as receptor expression (perhaps through classic ‘organizational’ effects in the brain during development; Hodgson et al. 2007). It is clearly erroneous to argue that hormone titles are more or less important than receptor levels or signal transduction; hormonal regulation requires a complex, integrated endocrine system with multiple levels of control. Although analytical and experimental utility will often dictate a focus on a single component, it is likely that marked inter-individual variation will be found at other levels of endocrine regulation, e.g. receptors, intracellular signalling, gene expression (Whitehead & Crawford 2006; Ball & Balthazart 2008; Nininmaa & Waser 2007) which will need to be integrated with data on inter-individual variation in hormone signal. Nevertheless, hormone titles, which are more easily measurable in more systems than are receptors, provide abundant and functionally significant phenotypic variation with which to explore structure–function relationships and questions about the evolution of endocrine systems.

6. ARE HORMONE TITRES CORRELATED WITH PHENOTYPIC VARIATION IN HORMONE-DEPENDENT TRAITS?

Correlations among physiological traits, or between traits and hormone titles, can be a valuable initial approach for inferring functional or mechanistic relationships (Bennett 1987; Zera et al. 2007). Given a priori knowledge of the different components and interrelatedness of endocrine-regulated networks, is it a reasonable expectation to find positive correlations...
between hormone titres and phenotypic variation in hormone-dependent traits? For example, for the hypothalamic-pituitary-gonadal axis, is there a systematic relationship between the marked (10-fold) inter-individual variation in plasma oestrogen levels and variation in either the hormonal stimulus for E2 release (e.g. plasma LH levels) or variation in oestrogen-dependent traits such as yolk precursors levels, yolk or egg size? If so, does this ‘explain’ inter-individual variation in hormone titres? In contrast, if we find no such correlations does this suggest that inter-individual variation in hormone titres is real but functionally neutral or that variation is simply maintained by mutation-selection balance?

Few studies have analysed inter-individual variation in hormone titres in this way. In a recent study, Carlson et al. (2006) reported a positive association between plasma cortisol levels and pup-feeding rates in male meerkats (Suricata suricatta) after controlling for other non-hormonal factors known to affect offspring care. In the cooperatively breeding Florida scrub jay (Aphelocoma coerulescens coerulescens), plasma prolactin levels were positively correlated with nestling feeding rate in non-breeding helpers at the nest, but not in breeders (Schoech et al. 1996), and in wandering albatross (Diomedea exulans), plasma corticosterone was negatively correlated with foraging trip success (Angelier et al. 2007). Numerous studies have reported positive correlations between plasma thyroid hormone levels (T3/T4) and growth in fishes (e.g. Gomez et al. 1997) or metabolism (Steyermark et al. 2005). These relationships are all consistent with putative or known physiological functions of these respective hormones, although correlations between hormone titres and phenotype are generally weak, i.e. there is substantial unresolved variation. Negative results are, of course, less often reported, but as an example, during egg production in birds, plasma oestradiol varies 10-fold between 44 and 423 pg ml\(^{-1}\) in females with complete follicle hierarchies (four or more yolky follicles). However, variation in two putative E2-dependent traits, plasma yolk precursor levels and total mass of yolky follicles, is independent of plasma E2 levels (even though yolk precursor levels themselves vary 10-fold; Williams et al. 2004). This lack of covariation in phenotypic variation in non-manipulated individuals contrasts with the clear dose-dependent response of plasma yolk precursor levels to exogenous hormone in this system (Williams & Martyniuk 2000), a contrast that is common for behavioural traits (Adkins-Regan 2005).

Ultimately, we are most interested not simply in covariation between hormone titres and phenotypic variation but whether hormonal variation affects fitness and thus how selection might have shaped the normal distribution of plasma T levels observed in natural populations. Compared with control males, high-T males have higher song rates, larger territory size, are more attractive to females and gain more extra-pair fertilizations, but they also show decreased parental behaviours (less nest defence, lower chick-feeding rates) and lower survival, perhaps related to a lower body fat, higher plasma corticosterone, suppressed immune function and delayed moult (Ketterson et al. 2001). Perhaps surprisingly, Reed et al. (2006) showed that high-T males (the ‘extreme’ phenotype) had higher fitness than control males, though solely due to higher rates of extra-pair copulation by high T-males, suggesting, in the absence of a comparable natural phenotype, that T levels are constrained in natural populations (Reed et al. 2006). These studies, and others using hormone manipulations (e.g. Sinervo 1999), have undoubtedly advanced our understanding of the endocrine basis of life histories. However, a major unanswered question generated by these studies is the extent to which experimental manipulation and the performance of extreme hormonal phenotypes mimics the pleiotropic effects of normal variation in endocrine regulators in non-manipulated individuals. Thus, the relevance of the manipulation studies to both the existence and fitness effects of natural endocrine genetic variation remains to be established (McGlothlin & Ketterson 2008; Zera et al. 2007).

Finally, for hormonal variation to be related to fitness and for endocrine mechanisms to respond to selection inter-individual variation must have a genetic basis, i.e. hormone titres or perhaps more importantly plasticity in hormone titres (e.g. the slopes in hormonal reaction norms) must be heritable. Heritability estimates are available for a range of absolute hormone titres and binding globulins, though mainly from twin studies in humans or for captive, laboratory-bred populations (Zera et al. 2007), and these studies demonstrate genetic variation and genetic correlations for various hormones with realized heritabilities generally less than 0.30 (e.g. Odeh et al. 2003; Evans et al. 2006). However, these studies also confirm that heritability estimates can vary markedly for different hormones and for the same hormone in different populations (e.g. comparing adolescent, middle-aged and elderly humans: Ring et al. 2005), i.e. heritability is a characteristic not only of a trait but also of a specific population (Falconer & McKay 1996). The data available to date are therefore of limited utility in understanding heritability of hormone titres in free-living vertebrates; indeed there appears to have been only one such study: King et al. (2004) reported heritabilities close to 1.0 for plasma T in male garter snakes (Thamnophis sirtalis; although the authors cautioned that these estimates might be inflated by maternal effects).

7. WHY DO SOME INDIVIDUALS HAVE MUCH HIGHER HORMONE TITRES THAN OTHER INDIVIDUALS?

It is clear that there is marked inter-individual variation in hormone titres but even where hormone titre–function relationships have been described,
correlations are often weak, i.e. there is large, unexplained residual variance. In some systems (e.g. E2-dependent yolk formation), individuals with very high circulating hormone levels appear to derive no functional benefit in terms of increased expression of hormone-dependent traits. Why then do some individuals maintain much higher hormone titres than other individuals? Here, I suggest three possible, non-mutually exclusive, explanations that deserve further consideration.

(a) Hormone titres within a cost–benefit framework

‘Direct’ costs of hormone production, e.g. the energy cost of biosynthesis, are generally thought to be small and inconsequential to the evolution of hormonal variation (though this appears never to have been quantified). However, hormones can have both beneficial (positive) or costly (negative) pleiotropic effects and individual hormone titres should reflect a trade-off between costs and benefits of these multiple physiological effects. For example, although corticosterone is essential in regulating routine metabolism, energy management and adaptive responses to acute stressors, even moderate chronic elevation of corticosterone can have negative effects on growth and immune function (Charmandari et al. 2005). Much attention has focused on the role of T in mediating various trade-offs based on the pleiotropic costs and benefits of this hormone. For example, T is thought to mediate a trade-off between mating effort and parental care (Wingfield et al. 1990), and this central concept has survived relatively well in the face of experimental study (at least in birds; Hirschenhauser et al. 2003; Lynn et al. 2005). In contrast, it has also been proposed that T mediates a trade-off between expression of sexual signalling traits and immune function (Folstad & Karter 1992): full signal expression requires high levels of T but this carries a cost due to the pleiotropic, immunosuppressive effects of T. Here, despite a very large number of experimental studies, there is at best only equivocal support for the central assumption of this trade-off: that T is immunosuppressive (Roberts et al. 2004). Furthermore, even post hoc modifications of this hypothesis, e.g. that T interacts with corticosterone to mediated the trade-off between signalling and immune function, have produced inconsistent or contradictory results (Roberts et al. 2007). Thus, attempts to understand variation in hormone levels in a cost–benefit framework have been limited to one or a few hormones, and they have met with mixed success, but from the perspective of this paper they have so far been restricted to interspecific differences.

Within such a cost–benefit framework, large-scale inter-individual variation presents a further paradox. If there are costs of high hormone levels, selection should generate a match between physiological capacity (hormone level) and functional demand (the amount of hormone required for physiological function; sensu Diamond & Hammond 1992) and this should reduce inter-individual variation. A possible explanation for this paradox is that individuals have different sensitivities to specific circulating hormone levels, such that in different individuals very different hormone titres are required to support the same level of physiological function. The idea that individuals maintain individually ‘optimized’ hormone titres, balancing relative costs and benefits at an individual level, has received little or no attention to date. Peters (2000) proposed that individually variable functional hormone titres, or differential sensitivities, might explain the positive correlation between natural variation in plasma T and immune function, but a negative effect of experimental treatment with T on immune function. One complication in addressing this issue is that we really need to measure the many combined or net physiological effects of a particular hormone. This argues strongly for an integrated approach to endocrine regulation, with measurement of multiple hormone-dependent outputs (e.g. Ketterson et al. 2001), rather than measurement of one or a few traits as is the case in many current studies (a situation analogous to the more recent recognition of the complexity of the immune system and the need to measure multiple traits; Viney et al. 2005).

(b) Evolutionary design: do individuals vary in ‘reserve capacity’?

In many physiological systems, individuals maintain higher functional capacities than required for ‘normal’ functional demand and this excess capacity over ‘load’ is referred to as reserve capacity or a ‘safety factor’ (Diamond 2002). Does this concept of reserve capacity explain the marked inter-individual variation observed for hormone titres? Biological safety factors vary from 1.25 to 10 for various structural components (e.g. bone morphology), measures of organ performance (filtration, absorption or secretion rate) or enzyme activity (Diamond 2002). However, this conceptual approach does not appear to have been applied to variation in hormone titres or other humoral components (although certain traits such as milk secretion from mammary glands and enzyme secretion from pancreas, with safety factors of 3 and 10, respectively (Diamond 2002) are hormone-dependent traits). It seems plausible that some individuals maintain high reserve-capacity hormone levels (i.e. higher plasma levels) and thus could obtain fitness benefits by more rapidly, or more effectively, matching demand during periods of upregulation of endocrine activity. Interestingly, endocrinologists have typically assumed that one function of binding globulins is to provide a reservoir or reserve of hormone (Norris 1997; see Landys et al. 2006) for another perspective on reserve capacity). Typically, 95–99% of circulating thyroid hormones, glucocorticoids and T are bound and (in theory) biologically inactive in this form. However, even in physiological states where the ‘demand’ for bioactive hormone presumably increases there is very little change in percentage bound hormone. For example, in alternate male phenotypes of tree lizards, which show differences in T-dependent aggressive behaviour, both territorial and non-territorial morphs have 99.7 and 99.8% of circulating T bound to binding globulins, respectively (Jennings et al. 2000). Furthermore, even with this high level of T binding only a small fraction (7%) of the total binding globulin capacity is taken up by T. Similarly, during the adrenocortical stress response, although there are marked changes in total and free circulating glucocorticoids (e.g. Wada et al. 2007) and corticosterone-binding globulin (CBG; Breuner et al. 2007).
Groothuis tages that later-hatched chicks face due to sibling later-laid eggs might ‘compensate’ for the disadvantage derived yolk hormones (e.g. from 5 to 20 pg mg$^{-1}$ for a-eggs comparing individuals 1 and 2 in figure 4). How is it that this much larger inter-individual variation does not generate much greater phenotypic variation among chicks of different females? why do not females producing eggs with such high yolk T levels give rise to ‘super-chicks’ compared with females producing eggs with very low T, which might be expected to be ‘super-duds’?

This issue of the functional significance of absolute hormone titres among individuals or the relative change in hormone titres within an individual has important implications for experimental studies where individuals are exposed to a standard hormone manipulation. As a result, a ‘standard’ manipulation will modify the hormone titres very differently for individuals with naturally low or high endogenous hormone levels. For example, ‘high’ hormone individuals might show a reduced response to manipulation if they are already close to some functional maximum hormone titre. Some support for the idea that this might be a more general issue comes from studies with oestriol manipulations, e.g. treatment of laying zebra finches with exogenous oestradiol (i.e. increasing the hormonal signal) has no effect on putative oestrogen-dependent traits such as yolk precursor levels, egg size or hematocrit. In contrast, reducing or blocking the hormonal signal (in this case using the anti-oestrogen tamoxifen) confirms E2-dependent effects on these same traits (Williams 2001; Wagner & Williams 2007). Negative effects of hormone treatment on phenotype should not therefore be taken as evidence that a particular hormone does not regulate phenotype, and future studies should not only augment but also block the hormone signal of interest.

8. CONCLUSIONS

Comparative and ecological endocrinologists continue to under-use the resource of inter-individual variation. In advancing the integration of endocrinology with evolutionary biology, especially in the context of animal responses to environmental change, we need to develop a much better understanding of the magnitude, mechanisms and functional significance of phenotypic variation, plasticity and flexibility in endocrine systems. A renewed focus on inter-individual variability provides both challenges to conventional wisdom in endocrinology itself, i.e. the way we currently view endocrine systems, and tremendous opportunities for endocrinologists to contribute significantly to evolutionary biology. Challenges include relatively simple, but unresolved questions such as structure–function relationships among hormones, binding globulins and receptors and the functional significance of absolute versus relative hormone titres. Opportunities abound for endocrinologists to provide a solid mechanistic understanding to key questions in evolutionary biology, e.g. how endocrine regulation is involved in evolution of complex suites of traits (such as those underlying concepts of ‘behavioural syndromes’ or ‘temperament’; Sih et al. 2004a; Reale et al. 2007), or how hormone

---

(c) **Is it absolute or relative hormone titre that is functionally important?**

Whether individuals ‘optimize’ circulating levels of hormones, or have individually variable sensitivity to a given hormone titre, this suggests that it is not the absolute value of the hormone titre but rather the relative hormone level, i.e. variation about an individual’s average or baseline level, that is important for determining phenotypic trait values. This issue has been largely ignored in many endocrine studies where functional, ‘adaptive’ explanations are provided for average or population-level changes in hormone levels. This is illustrated in figure 4. In many avian species concentrations of maternally derived yolk hormones vary systematically with laying order of the egg. It has been hypothesized that higher yolk androgen levels in later-laid eggs might ‘compensate’ for the disadvantages that later-hatched chicks face due to sibling competition with larger, earlier-hatched chicks (e.g. Groothuis et al. 2006). This might arise, for example, through a positive effect of higher yolk T on begging rate and post-natal growth (Eising & Groothuis 2003). However, from figure 4, this adaptive adjustment of yolk hormone levels occurs over a mean range of 14–18 pg mg$^{-1}$ comparing a- and c-eggs, i.e. this hypothesis suggests variation of less than 5 pg mg$^{-1}$ yolk T is sufficient to drive phenotypic differences between early- and late-hatching chicks. However, for any given egg sequence, inter-individual variation in yolk T is between three- and fourfold greater than this (e.g. a-egg b-egg c-egg).

---

**Figure 4.** ‘Adapative’ hypotheses which propose effects of variation in mean hormone levels on phenotypic traits fail to explain why much larger inter-individual variation does not generate even greater phenotypic variation; this is illustrated with data on laying sequence-specific variation in yolk androgen levels (based on data in fig. 1d, Groothuis et al. 2006; see text for more details).

Phil. Trans. R. Soc. B (2008)
pleiotropy regulates trade-offs among life-history traits. This will require endocrinologists to embrace the raw material of adaptation: heritable, individual variation or phenotypic plasticity, and to take advantage of conceptual approaches widely used in evolutionary biology, e.g. selection studies, reaction norms, concepts of evolutionary design. Finally, if we wish to integrate endocrinology and evolutionary biology, future progress will require that endocrinologists focus more explicitly on the life-history traits that are of primary interest to evolutionary biologists: fecundity (egg size and number), variation in breeding schedules, and survival (cf. behavioural endocrinology), key traits for which we still have an alarmingly poor understanding of mechanisms generating and maintaining phenotypic variation (Williams 2005; Sockman et al. 2006).

I would like to thank Kathy Wynne-Edwards, David Green, Julian Christians, Denis Reale, Marcel Lambrechts, John Wingfield, Tony Zera and two anonymous referees for their enlightening discussion and comments on drafts of this paper. Thanks also to my graduate students, especially Oliver Love, Katrina Salvanate, Emily Wagner and Francois Vezina for putting up with my rants on ‘variability’ while giving me invaluable comments and perspective on the ideas in this paper. Emily Wagner and Katrina Salvanate provided data for figure 3. This work was supported by an NSERC Discovery grant which supported Canadian participation in E-Bird project. Emily Wagner and Katrina Salvante provided data for the figure 3. This work was supported by an NSERC Discovery grant to T.D.W. and an NSERC Special Research Opportunity grant which supported Canadian participation in E-Bird and facilitated the stimulating, interdisciplinary discussion at the various workshops and meetings.

REFERENCES


