How can we estimate natural selection on endocrine traits? Lessons from evolutionary biology

Frances Bonier and Paul R. Martin

Biology Department, Queen’s University, Kingston, Ontario, Canada K7L 3N6

An evolutionary perspective can enrich almost any endeavour in biology, providing a deeper understanding of the variation we see in nature. To this end, evolutionary endocrinologists seek to describe the fitness consequences of variation in endocrine traits. Much of the recent work in our field, however, follows a flawed approach to the study of how selection shapes endocrine traits. Briefly, this approach relies on among-individual correlations between endocrine phenotypes (often circulating hormone levels) and fitness metrics to estimate selection on those endocrine traits. Adaptive plasticity in both endocrine and fitness-related traits can drive these correlations, generating patterns that do not accurately reflect natural selection. We illustrate why this approach to studying selection on endocrine traits is problematic, referring to work from evolutionary biologists who, decades ago, described this problem as it relates to a variety of other plastic traits. We extend these arguments to evolutionary endocrinology, where the likelihood that this flaw generates bias in estimates of selection is unusually high due to the exceptional responsiveness of hormones to environmental conditions, and their function to induce adaptive life-history responses to environmental variation. We end with a review of productive approaches for investigating the fitness consequences of variation in endocrine traits that we expect will generate exciting advances in our understanding of endocrine system evolution.

1. Introduction

Evolutionary endocrinologists seek to understand how endocrine traits are shaped by selection to form the patterns of trait variation evident among species, populations and individuals [1,2]. When considering among-individual variation—the variation upon which natural selection acts—evolutionary endocrinologists have drawn on classic approaches from evolutionary biology, particularly correlations between trait values (phenotypes) and fitness, to estimate selection (e.g. [3,4]). This approach, however, is likely to generate misleading and inaccurate estimates of both natural selection and the predicted responses to selection for traits that exhibit adaptive plasticity [5–11]. The goal of this paper is to illustrate the limitations associated with using phenotype–fitness correlations for studying selection on endocrine traits and to review more robust approaches, drawing from the work of evolutionary biologists who recognized this issue as it applied to other plastic traits decades ago.

For clarity, here we define plasticity in the broadest terms: the ability of one genotype (or individual) to express multiple phenotypes, which change in response to environmental conditions [12,13]. Thus, the term *phenotypic plasticity* encompasses both reversible plasticity—as is seen with changes in circulating hormone concentrations in response to short-term changes in the environment (sometimes referred to as *flexibility* in the endocrine literature [14,15])—as well as developmental and other forms of more stable and irreversible plasticity—as is seen in organizational effects of early environmental conditions, which can have lasting effects on the sensitivity and activity of a number of endocrine axes...
among-individual variation observed in the trait in the population and the contribution of additive genetic variation to the total of among-individual variation in the trait in the population (diction follows from the ‘breeder’s equation’, where the component \[20–24\]). These observations lead to the prediction suggests that variation in breeding date has a heritable component. Further, evidence from several populations of birds commonly developed a theory to illustrate the problem. In this classic example, the earliest-breeding individuals within a population often have the highest fitness, and yet the most common breeding date is later than this apparent optimum (grey bars), and is associated with lower fitness. This relationship suggests directional selection for earlier breeding date; however, the predicted evolutionary response to selection does not occur. The lack of an evolutionary response to selection on breeding date is probably driven by the absence of directional selection; the effects of nutritional state on both breeding date and fitness (both phenotypically plastic traits) could create the observed relationships as proposed by Price et al. [8] and others [12,13].

[16,17]. This broad definition follows usage in evolutionary biology and other fields (reviewed in [12]).

When traits exhibit plasticity and adaptively respond to environmental factors that also influence fitness components, estimating selection using phenotype–fitness correlations becomes problematic [5,6,9,10]. Darwin [18] and Fisher [19] both alluded to this problem as it relates to variation in the timing of breeding, and later Price and colleagues [8] explicitly developed a theory to illustrate the problem. In this classic example, the earliest-breeding individuals within a population often have the highest fitness, and yet the most common breeding date is later than this apparent optimum (figure 1), suggesting directional selection for earlier breeding. Further, evidence from several populations of birds suggests that variation in breeding date has a heritable component [20–24]. These observations lead to the prediction that breeding date should evolve to become earlier. This prediction follows from the ‘breeder’s equation’, where the population-level response to selection between generations (R) is predicted by the product of the narrow sense heritability of among-individual variation in the trait in the population (\( h^2 \); or the contribution of additive genetic variation to the total among-individual variation observed in the trait in the population) and the strength of selection (\( s \); proportional to the difference between the optimal trait value and the mean observed population trait value, or the correlation between individual phenotypic variation in the trait and fitness metrics for those individuals; i.e. the phenotype–fitness correlation). The modal breeding date, however, does not shift over time to approach the optimum, suggesting a failure of breeding date to respond to directional selection [25]. This apparent paradox can be explained by variation in the nutritional state of individuals that influences both the timing of breeding and, independently, fitness components [8]. The individuals with the highest nutritional state are able to breed earlier and are also able to produce larger clutches than individuals in a poorer condition [26–28] (but see [29]), generating a positive association between breeding date and fitness driven by plastic responses of both traits to nutritional state, and incorrectly suggesting directional selection on breeding date. Instead, the association between breeding date and fitness probably represents stabilizing selection, favouring individuals that optimize their breeding date according to their nutritional condition, and thus the negative correlation between breeding date and fitness is maintained over time [8]. In other words, selection probably favours heritable variation for both plasticity in breeding date and plasticity in reproductive investment (i.e. reaction norms) that allow individuals to adaptively shift their reproductive timing and investment in response to changing nutritional conditions.

Using phenotype–fitness correlations to estimate selection on plastic traits, like optimal breeding date, can lead to incorrect estimates of the shape and even direction of the selection gradient [9]. Moreover, the limitations of this method apply even more strongly to endocrine traits than they do to many other plastic traits because of the dynamic plasticity of hormones in response to environmental conditions, and their function to induce adaptive life-history responses of the organism to variable environmental conditions. In other words, not only might endocrine traits change in response to environmental factors that also influence fitness, these changes in hormones in some cases are the very signals that induce adaptive shifts in investment in reproduction or self-maintenance, thereby driving individual variation in fitness. Thus, we expect that correlations between endocrine traits and fitness components should be widespread because of their closely linked patterns of adaptive plasticity, independent of any correlation reflecting natural selection. Yet, across the evolutionary endocrinology literature, the most common manner for estimating how contemporary selection is acting on individual variation in endocrine traits is inference from phenotype–fitness correlations (e.g. [3,4,30–35]).

2. An example from evolutionary endocrinology

To illustrate this problem as it applies to evolutionary endocrinology, we use the example of circulating glucocorticoid concentrations and nutritional state below. We note, however, that the same issues will apply to any endocrine trait (e.g. other hormones, hormone receptors, binding globulins, and even the magnitude or slope of endocrine responses) when trait values respond plastically to any environmental factors that also influence fitness.

In a population of tree swallows in Ontario, Canada, that we have studied for the past 10 years, baseline corticosterone (CORT, the primary glucocorticoid in birds) concentrations in breeding females can be negatively correlated with fledging success (the number of offspring successfully raised to the age of departing the nest), yet the most common CORT values are higher than the apparent optimum (figure 2). Several studies have estimated non-zero heritability of individual variation in circulating CORT concentrations, through artificial selection [36,37], estimates of repeatability [38–40] and animal model approaches [41]. Thus, upon characterizing a correlation between CORT and a fitness proxy, and with the
expectation of some degree of heritability underlying observed among-individual variation in CORT, we might describe figure 2 as evidence for directional selection for lower CORT levels, and predict the evolution of reduced CORT levels over time. These inferences—and others that we could make about the magnitude, shape and even direction of selection—are most probably incorrect for the same reasons as those described in the breeding date example above.

The problem with using a phenotype–fitness correlation to infer selection on CORT again lies with the plasticity of individuals. For simplicity, we can consider the same environmental factor and fitness proxy in this example as described in relation to breeding date: nutritional state and clutch size. Within one individual, CORT concentrations are plastic, typically increasing in response to declining nutritional state (figure 3a) [42–44]. This within-individual plasticity of CORT has probably been shaped by selection; for example, increasing CORT in response to declining nutritional state can adaptively induce increased foraging behaviour [42,45,46]. The same individual with a low nutritional state will also produce the smallest clutches of eggs because she has fewer resources to invest in reproduction (figure 3b). Again, this plasticity of clutch size has been shaped by selection—optimal reproductive investment is tightly linked to environmental conditions, including nutritional state [47,48]. Furthermore, the changing levels of reproductive investment, as reflected in clutch size in this example, might even be induced by changing levels of CORT [49,50]. Although natural selection might favour plastic responses of both CORT and clutch size to variation in nutritional state, these parallel, adaptive responses generate a within-individual correlation between CORT and fitness that superficially suggests directional selection (figure 3c). If we had characterized one individual’s CORT and clutch size across a number of breeding bouts without information on nutritional state, we might conclude that the optimal CORT level for this individual is low, because her maximal fitness coincides with a low CORT level; in reality, her optimal CORT level and clutch size varies depending on her nutritional state.

If we characterize the plastic responses of both CORT and clutch size to varying nutritional states across a number of individuals within a population, we again should expect CORT to increase (figure 4a) and clutch size to decrease (figure 4b), as the nutritional state of any individual declines. At any given sampling time, individuals are probably experiencing different nutritional states due to variation in resource availability, past and current energetic demands, and ability to incorporate and allocate resources. Thus, if we estimate CORT levels and clutch sizes for a random sample of individuals across a population, we would predict a negative CORT–clutch size relation that is driven entirely by the adaptive plastic responses of individuals to their different nutritional states. This phenotype–fitness correlation might not reflect the shape or even the direction of natural selection on CORT (figure 4c). For example, this phenotype–fitness correlation is consistent with directional selection, stabilizing selection and no selection on CORT levels. As with the example of timing of breeding, the problems with using phenotype–fitness correlations to infer selection does not mean that selection is not acting on CORT; instead, it illustrates that hormone–fitness correlations cannot reveal if or how selection is acting.

### 3. Further complications for estimating selection from phenotype–fitness correlations

The example above uses nutritional state to illustrate the problem of estimating selection on endocrine traits using phenotype–fitness correlations, and might suggest that we could circumvent this limitation if we controlled for nutritional state. However, controlling for nutritional state alone cannot address the broader issue because a wide array of other environmental factors contribute to the same problem. For example, variations in weather [51–53], conspecific density [54–56], developmental conditions [57–59], predation risk [60–64], disease [65–68] and residual reproductive value [69–72] are all thought to induce adaptive plasticity in both endocrine and fitness-related traits, thus confounding the use of phenotype–fitness correlations for inferring patterns of selection. The diversity of environmental factors influencing both endocrine traits and fitness makes it unfeasible to statistically control for these factors in studies of natural populations. Moreover, environmental variation can induce plastic responses with lasting effects across developmental stages [73–75] and even generations [54,76,77], further complicating any effort to control for environmental factors within a phenotype–fitness correlation framework.

Endocrine traits such as CORT that are highly responsive to the same environmental conditions that influence fitness (e.g. [51,52,54,60,61,64]) should be most prone to environmentally driven correlations between phenotypes and fitness. Other endocrine traits, however, will also suffer from the same concerns. For example, hormone binding globulins (e.g. [78]), reproductive steroids (e.g. [79]), protein hormones (e.g. [80]) and hormone receptors (e.g. [81]) can all change in response to environmental conditions that also influence fitness. Some evolutionary endocrinologists have measured how fitness relates to induced endocrine responses (or endocrine reaction norms), such as the elevation of testosterone following injection of a releasing hormone (e.g. [31]). The responsiveness and capacity of endocrine axes might be less labile than baseline circulating concentrations [38,40], and so these phenotype–fitness correlations might be less influenced by environmental bias and thus be somewhat more reliable for estimating selection. However, the responsiveness and
Below, we suggest means of achieving this goal. To advance our field, we need an understanding of the way measure of circulating hormone concentrations [1,84,85] to match its current conditions, rather than any single point to adaptively regulate and adjust its endocrine phenotype. For endocrine traits, the most likely target for selection is an individual’s ability shaped plasticity of the trait of interest. For endocrine traits, this approach would not directly address how selection has is challenging in natural populations, however, and this approach will be a productive route for advancing evolutionary endocrinology [86]. This approach presents greater methodological challenges than the endocrine phenotype–fitness correlation method, but should produce clearer and more valuable results.

To minimize the bias associated with environmentally induced plasticity in estimates of selection, one could derive correlations between genetic variation for both fitness and endocrine phenotypes [9,10,25]. Determining these relations is challenging in natural populations, however, and this approach would not directly address how selection has shaped plasticity of the trait of interest. For endocrine traits, the most likely target for selection is an individual’s ability to adaptively regulate and adjust its endocrine phenotype to match its current conditions, rather than any single point measure of circulating hormone concentrations [1,84,85]. To advance our field, we need an understanding of the way that selection has shaped the plasticity of endocrine traits [15]. Below, we suggest means of achieving this goal.

4. Moving forward: approaches to the study of selection on endocrine traits

Given the limitations of the phenotype–fitness correlation approach to estimating selection on endocrine traits, and the likelihood that the genetic basis for plasticity and regulation of these traits is a target of selection, a reaction norm approach will be a productive route for advancing evolutionary endocrinology [86]. This approach presents greater methodological challenges than the endocrine phenotype–fitness correlation method, but should produce clearer and more valuable results.

The estimation of a reaction norm—or the pattern of phenotypic expression of a single genotype over a range of environments [87]—has not been widely applied in endocrinology [85,88]. However, we routinely measure within-individual changes in hormone concentrations and other endocrine traits across life-history stages, environments and experimental treatments, which could be considered reaction norms of hormone titres to changing conditions (e.g. [3,89,90]). Less commonly, we also estimate repeatability of individual endocrine responses, though most often to artificial challenges like capture and restraint stress (e.g. [38,40,91], but see [92]). To our knowledge, no study has estimated both the endocrine and fitness responses of individuals to a range of ecologically relevant environments, nor the heritability of those reaction norms.
To effectively employ a reaction norm approach to estimate how selection acts on the plasticity of endocrine traits requires quantifying both the endocrine trait of interest and a fitness metric of individuals (or genotypes) randomly exposed to a range of experimentally manipulated conditions (figure 5) [93]. The endocrine reaction norm that produces the highest relative fitness across all environments could be inferred to be favoured by selection, although the frequency of occurrence of different environments in nature could lead to selection favouring reaction norms that have the highest fitness in the most common environments [94,95], or alternative reaction norms adapted to different environments (e.g. if there are trade-offs to maximizing fitness in different environments). To further predict future responses to selection, one would need to estimate the relation between heritable variation in both the reaction norms and fitness, which could be accomplished through pedigrees or selection experiments.

Given the need to repeatedly measure both fitness and endocrine phenotypes within individuals across multiple environments, iteroparous organisms that are amenable to repeated sampling and breeding in a mesocosm or semi-natural enclosure might be the most productive systems for studies of selection on endocrine trait reaction norms. Ideally, individuals derived from a natural population would be housed in a common garden for captive breeding to produce a common F1 or F2 stock for subsequent random assignment to experimental treatments. The use of wild individuals could help to avoid the problem of reduced variation common in domestic stocks of organisms, while the production of F1 and F2 stock can help to control for maternal and early environmental effects that could introduce bias into reaction norm–fitness correlations.

The reaction norm approach could be used to address central questions in evolutionary endocrinology, providing tangible advances in our field. For example, this approach could reveal alternate endocrine strategies among individuals adapted to different environmental conditions or with different life-history strategies. Reaction norm approaches could also be used to test hypotheses about single or interacting selective pressures operating in natural populations by exposing individuals to experimentally manipulated conditions that fall within the range of conditions experienced in nature. For predictive studies aimed at determining how populations might respond to changing conditions (e.g. climate change), exposure to conditions in the range of those expected to exist in the future would be informative. Complementary approaches could also be used over several generations to measure evolutionary responses of endocrine reaction norms to selection associated with experimental environments.

Several other classical evolutionary approaches could also be productive avenues for understanding how selection shapes endocrine traits. For example, artificial selection on endocrine, life-history and/or behavioural traits to create selected lines can help to elucidate the role of hormones in regulating fitness components (e.g. [96–99]). Selected lines can also produce individuals of known, genetically based endocrine phenotypes for exposure to natural selection in experimental settings (e.g. [97,98]). Organisms with genetically based life-history polymorphisms also provide excellent opportunities to test evolutionary hypotheses in individuals with known, genetically based variation in life-history strategies (e.g. side-blotched lizards [100], crickets [96] and white-throated sparrows [101]). Importantly, studies of polymorphic species could be conducted in the field on free-ranging individuals, provided the difference between morphs is the primary focus (i.e. the differences in apparent selection associated with known genetic differences).

The increasing pool of published data on endocrine traits also provides a rich resource for meta-analyses and comparative studies that can reveal broader-scale evolutionary patterns [69,102,103]. These comparative findings cannot reveal contemporary selection acting within populations, but can identify the selective pressures associated with the evolution of different endocrine traits or strategies across species or populations. Finally, correlations between hormones and fitness components, although not useful for estimating natural selection, can generate experimentally testable hypotheses regarding the functional role of endocrine plasticity in regulating traits that influence fitness (e.g. [53,54,89,104]). With predictions in hand from comparative studies or from within-population hormone–fitness correlations, field experiments can provide effective tests for contemporary selection. For
example, with endocrine traits that can be blocked or inhibited, one could hinder what is hypothesized to be an adaptive endocrine response, while simultaneously exposing individuals to an experimentally altered environment thought to favour that plastic response, and measure resulting changes in fitness relative to controls with intact endocrine traits.

5. Conclusion
Evolutionary endocrinology has generated novel insights and advances in our understanding of how selection shapes endocrine traits (e.g. [54,69,96,97,100,105]). Our progress, however, has been hindered by the widespread application of phenotype–fitness correlations that cannot accurately estimate selection on endocrine traits. We recommend that endocrine phenotype–fitness correlations no longer be used to estimate contemporary natural selection. Instead, reaction norms, selected lines, organisms with polymorphic life-history strategies, comparative studies and field experiments offer exciting and potentially productive routes forward.

Data accessibility. The tree swallow data used to generate figures 1 and 2 are available as electronic supplemental material.

Authors’ contributions. F.B. conceived of the ideas and drafted the manuscript; P.R.M. contributed to the ideas and helped draft the manuscript.

Competing interests. We have no competing interests.

Funding. Funding from the US National Science Foundation (IOS #1145625 to F.B. and I. T. Moore), Natural Sciences and Engineering Research Council of Canada, and Canadian Foundation for Innovation supports the ongoing tree swallow study.

Acknowledgements. We would like to thank E. Adkins-Regan, R. Cox, M. Hau, A. Lendvai, J. McGeilhlin, I. Moore, J. Ouyang and three anonymous referees for critical feedback and discussions that strengthened the final version of this manuscript.

References

32. McGeilhlin JW, Whittaker DJ, Schrock SE, Gerlach NM, Jawor JM, Snajdor EA, Ketterson ED. 2010
Natural selection on testosterone production in a wild songbird population. *Am. Nat.* 175, 687 – 701. (doi:10.1086/526469)


61. Ardia DR. 2005 Tree swallows trade off immune function and reproductive effort differently across
81. Dickens M, Romero L, Cyr N, Dunn I, Meddle S. 2009 Chronic stress alters gluocorticoid receptor and mineralocorticoid receptor mRNA expression in the European starling (Sturnus vulgaris) brain.


