

# Environmental Change, If Unaccounted, Prevents Detection of Cryptic Evolution in a Wild Population

Tomos Potter,<sup>1,\*</sup> Ronald D. Bassar,<sup>2</sup> Paul Bentzen,<sup>3</sup> Emily W. Ruell,<sup>4</sup> Julián Torres-Dowdall,<sup>4,†</sup> Corey A. Handelsman,<sup>4</sup> Cameron K. Ghalambor,<sup>4,5</sup> Joseph Travis,<sup>5</sup> David N. Reznick,<sup>6</sup> and Tim Coulson<sup>1</sup>

1. Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom; 2. Department of Biology, Williams College, Williamstown, Massachusetts 01267; 3. Department of Biology, Dalhousie University, Halifax, Nova Scotia BH3 4R2, Canada; 4. Department of Biology, Center for Biodiversity Dynamics, Norwegian University of Science and Technology (NTNU), Trondheim, Norway; 5. Department of Biological Science, Florida State University, Tallahassee, Florida 32306; 6. Department of Evolution, Ecology and Organismal Biology, University of California, Riverside, California 92521

Submitted October 4, 2019; Accepted July 7, 2020; Electronically published November 25, 2020

Online enhancements: supplemental PDF. Dryad data: <https://doi.org/10.5061/dryad.jm63xsj7k>.

**ABSTRACT:** Detecting contemporary evolution requires demonstrating that genetic change has occurred. Mixed effects models allow estimation of quantitative genetic parameters and are widely used to study evolution in wild populations. However, predictions of evolution based on these parameters frequently fail to match observations. Here, we applied three commonly used quantitative genetic approaches to predict the evolution of size at maturity in a wild population of Trinidadian guppies. Crucially, we tested our predictions against evolutionary change observed in common-garden experiments performed on samples from the same population. We show that standard quantitative genetic models underestimated or failed to detect the cryptic evolution of this trait as demonstrated by the common-garden experiments. The models failed because (1) size at maturity and fitness both decreased with increases in population density, (2) offspring experienced higher population densities than their parents, and (3) selection on size was strongest at high densities. When we accounted for environmental change, predictions better matched observations in the common-garden experiments, although substantial uncertainty remained. Our results demonstrate that predictions of evolution are unreliable if environmental change is not appropriately captured in models.

**Keywords:** animal model, breeder's equation, secondary theorem of selection, Robertson-Price identity, estimated breeding values, eco-evolutionary feedback.

## Introduction

Predictions of evolution based on quantitative genetic theory frequently fail to match observations of change (or lack thereof) in wild populations (Merilä et al. 2001; Gienapp et al. 2008; Pujol et al. 2018; Walsh and Lynch 2018). Detecting contemporary evolution in natural populations is challenging enough (Merilä et al. 2001; Kruuk 2004; Kruuk et al. 2008; Pemberton 2010); understanding why predictions about evolution often fail creates a much larger and more important challenge (Pujol et al. 2018).

There are three widely used quantitative genetic approaches for modeling contemporary evolution: (1) the breeder's equation, (2) the secondary theorem of selection, and (3) the change in estimated breeding values. The univariate breeder's equation (Lush 1937) predicts the change in the population mean trait value due to selection and is formalized as

$$\Delta z = h^2 S, \quad (1)$$

where  $\Delta z$  is the response to selection (i.e., the change in mean breeding value per generation),  $h^2$  is the (narrow-sense) heritability of the trait, and  $S$  is the selection differential. The heritability is the proportion of population-level variation in a trait that can be explained by genetic differences among individuals, given by the ratio of the additive genetic variance  $V_A(z)$  to the total phenotypic variance  $V_P(z)$ . The selection differential describes the relationship between the trait and fitness, which is given by their phenotypic covariance  $\text{cov}_z(z, w)$ . The breeder's equation in its multivariate form allows predictions of responses to selection for any number of genetically correlated traits (Lande and Arnold 1983). By accounting for the genetic covariance among multiple traits and their relationship

\* Corresponding author; email: [tomos.potter@zoo.ox.ac.uk](mailto:tomos.potter@zoo.ox.ac.uk).

† Present address: Department of Biology, University of Konstanz, Universitätsstrasse 10, 78457 Konstanz, Germany.

**ORCID:** Potter, <https://orcid.org/0000-0003-3201-6130>; Torres-Dowdall, <https://orcid.org/0000-0003-2729-6246>; Ghalambor, <https://orcid.org/0000-0003-2515-4981>; Travis, <https://orcid.org/0000-0002-2419-5020>; Reznick, <https://orcid.org/0000-0002-1144-0568>.

Am. Nat. 2021. Vol. 197, pp. 29–46. © 2020 by The University of Chicago. 0003-0147/2021/19701-59530\$15.00. All rights reserved.  
DOI: 10.1086/711874

to relative fitness, the multivariate breeder's equation should accurately predict evolutionary change, provided that all correlated traits that influence fitness are included in the model (Lande and Arnold 1983).

An alternative approach to describing evolutionary change was developed by Robertson (1966) and independently formalized by Price (1970) shortly afterward. The secondary theorem of natural selection (also known as the Robertson-Price identity) states that the expected change in mean trait value per generation,  $\Delta z$ , is equal to the additive genetic covariance  $\text{cov}_A$  of the trait  $z$  and relative fitness  $w$ :

$$\Delta z = \text{cov}_A(z, w). \quad (2)$$

By considering only the genetic relationship between a trait and relative fitness, the secondary theorem of selection does not require all traits that influence fitness to be measured, nor does it assume that selection on the phenotype is equal to selection on the genotype, as in the breeder's equation (Morrissey et al. 2010, 2012). As a result, predictions of evolutionary change from the breeder's equation and the secondary theorem of selection may differ, with the latter often considered the more robust model of evolutionary change (Morrissey et al. 2010, 2012; Bonnet et al. 2017).

A third method of quantifying contemporary evolution is to describe the change in estimated breeding values over time. An individual's breeding value is the sum of the additive genetic effects on its phenotype (Lynch and Walsh 1998). While a change in true breeding values over time is synonymous with evolutionary change, care is required when interpreting changes in estimated breeding values (Postma 2006; Hadfield 2008; Hadfield et al. 2010). Many applications of this approach in studies of wild populations have generated biased and anticonservative estimates of evolutionary change (reviewed and discussed in Hadfield et al. 2010). However, with appropriate modeling frameworks, contemporary evolution can be quantified by the temporal change in the population mean predicted breeding value.

All of these approaches rely on accurate decomposition of phenotypic variances and covariances into additive genetic and environmental components. The statistical foundation for this decomposition is the so-called animal model. Animal models are mixed effects models for the phenotype that are structured with a relatedness matrix (Henderson 1950; Lynch and Walsh 1998; Kruuk 2004). The basic animal model assumes an additive genotype-phenotype map, where an individual's phenotype  $P$  (with respect to the population mean value) is given by the sum of the breeding value  $G$  and the environmental component  $E$ —that is,  $P = G + E$ . Animal models estimate  $G$  by assessing phenotypic similarity among all pairs of rela-

tives, while  $E$  is described by the residual (Lynch and Walsh 1998). As such,  $E$  is broadly defined and captures not only the effect of environmental factors on the expression of traits (i.e., phenotypic plasticity) but also those of nonadditive genetic effects (i.e., dominance and epistasis; Lynch and Walsh 1998). Animal models maximize the information on relatedness obtained from natural pedigrees and are robust to (randomly) missing and unbalanced data, and so they are considered well suited to inferring quantitative genetic parameters in natural systems (Lynch and Walsh 1998; Kruuk 2004; Wilson et al. 2010). Animal models are frequently used to quantify evolution in the wild, by estimating breeding values and parameters for the breeder's equation and the secondary theorem of selection (Merilä et al. 2001; Grant and Grant 2002; Kruuk et al. 2008; Charmantier et al. 2014). However, the use of animal models in wild populations has revealed that despite the presence of heritable genetic variance and selection, there is frequently no observable change in the mean phenotype, a phenomenon termed “the paradox of stasis” (Merilä et al. 2001; Gienapp et al. 2008; Pujol et al. 2018; Walsh and Lynch 2018).

The crux of the paradox of stasis is that quantitative genetic approaches can accurately describe the response to selection in captive-bred organisms but not in wild populations. A logical first step in resolving this paradox is to identify fundamental differences between artificial and natural selection. The main difference, undoubtedly, is that breeders target selection on specific traits, whereas natural selection acts on the full phenotype of the organism. The multivariate breeder's equation addresses this issue to a degree by accounting for indirect selection through genetically correlated traits (Lande and Arnold 1983). The estimation of breeding values is also sensitive to the omission of correlated traits that are under selection; bias in estimation of breeding values is reduced when correlated traits are included in the animal model (Pollak et al. 1984). The secondary theorem of selection bypasses this problem by considering selection only at the genetic level and making no inference about selection at the phenotypic level (Robertson 1966; Price 1970; Morrissey et al. 2012).

Another critical difference between artificial and natural selection is that of environmental variability (Walsh and Lynch 2018, chap. 20). Animal and plant breeders purposefully reduce environmental variation among individuals by making rearing conditions as similar as possible and providing nutrition, shelter, and protection from disease. In contrast, environmental factors in the wild may change substantially over time and space, generating considerable variation in phenotype and fitness within the population. In addition, dynamic environments may result in fluctuations in the target, magnitude, and direction of selection (e.g., Price et al. 1984; Hairston and Dillon

1990; Grant and Grant 2002); ignoring such effects could lead to erroneous predictions of evolution in the wild.

Not only does environmental variability have a direct influence on evolution in wild populations, it can also affect the results of statistical approaches we use to quantify evolutionary change. Here, we consider how subgroup effects—environmental sources of phenotypic variation that differ in magnitude between subgroups of a population—can bias estimates of quantitative genetic parameters from animal models. Subgroup effects vary spatio-temporally, such that there are broad-scale differences in the environmental conditions experienced among groups of individuals within the population as a whole. If subgroup effects are not explicitly accounted for, two distinct problems may arise. First, when relatives experience different subgroup effects, estimation of additive genetic (co) variances will be biased. This effect can occur within a generation between siblings or cousins or between generations between parents and offspring. Second, subgroup effects can lead to the regression model not capturing the causal relationship between traits and fitness. Both of these problems may lead to misestimation of the evolutionary trajectory.

Regarding the first problem, if relatives consistently experience different subgroup effects, the genotype-phenotype map is no longer strictly additive, because  $G$  and  $E$  are not independent. Estimation of  $V_A$  assumes that phenotypic similarity among relatives is due to shared genetic elements of additive effect on the phenotype (Lynch and Walsh 1998). Nongenetic sources of similarity among relatives are routinely taken into consideration when estimating  $V_A$  using animal models (e.g., maternal effects [Wilson et al. 2010], birth year [Kruuk and Hadfield 2007], common rearing environment [Kruuk et al. 2001], and spatial autocorrelation [Stopher et al. 2012]). Doing so prevents upward bias when estimating  $V_A$ .

Nongenetic causes of dissimilarity (rather than similarity) among relatives are less often accounted for. If parents and offspring experience different environments, their phenotypes and fitness will be less similar than if they had experienced the same conditions. For example, in a population growing toward carrying capacity, parents will experience higher per capita resource availability than their offspring; that is, there is a difference in subgroup effects between parents and offspring with respect to resource availability. This environmentally induced dissimilarity among relatives could result in underestimation of  $V_A$  and misestimation of genetic covariances and breeding values.

When subgroup effects vary over time, a special case of the first problem occurs when estimating the change in breeding values. Estimation of breeding values in a simple animal model assumes that the environment is constant

in time and space across the whole population (Postma 2006; Mrode 2014). Under this assumption, estimated breeding values will partially reflect the environmental component of the phenotype (Postma 2006). As such, when subgroup effects are not modeled, a change in estimated breeding values may reflect phenotypic rather than true genetic change (Postma 2006; Hadfield et al. 2010).

Even when relatives experience the same subgroup effects, the second problem—of regression models capturing noncausal associations between traits and fitness—can arise when selection is estimated across the entire population, as is typical in studies of wild populations using animal models. Environmental factors that influence both traits and fitness can bias estimates of selection gradients (Rausser 1992; Stinchcombe et al. 2002). For example, consider a population in which two subgroups differ only in experiencing high or low resource availability. If trait  $z$  and fitness  $w$  are both greater at high resource availability, the phenotypic selection gradient will be positive when estimated across both subgroups even if  $\text{cov}_z(z, w)$  within subgroups is zero. In this case, the regression model captures a noncausal association between  $z$  and  $w$ ; the phenotypic selection gradient is entirely dependent on environmental factors rather than any real effect of the trait on fitness.

These statistical problems are likely inherent to studies of wild populations. A potential solution is to include subgroup-specific measurements of environmental factors as covariates in the animal model, to “correct” phenotypes and fitness prior to estimating breeding values, additive genetic variances, and selection. Including subgroup effects as covariates changes the mean part of the animal model. Instead of having a single population mean value around which  $G$  and  $E$  are estimated, the model fits subgroup-specific mean values, accounting for subgroup-specific differences in environmental factors. This reduces the problems of nonindependence of  $G$  and  $E$  and of noncausal associations between traits and fitness, as described in the previous paragraphs. However, for wild populations it can be difficult to identify and quantify the relevant environmental factors (Walsh and Lynch 2018). Nevertheless, not accounting for subgroup effects may give rise to incorrect estimates of quantitative genetic parameters in wild populations, precluding accurate estimation of evolutionary change.

Regardless of which model is used to predict phenotypic change, quantitative genetic predictions and phenotypic observations from field data are rarely corroborated with independent data characterizing evolutionary change of the same population (although see Van Asch et al. 2013; Geerts et al. 2015). Common-garden experiments, in which population samples from before and after selection are reared under identical environmental

conditions, can determine whether a change in the mean phenotype is due to evolution (Kawecki and Ebert 2004; Pemberton 2010). Common-garden experiments remove the influence of subgroup effects on the phenotype and thereby overcome the challenges posed by spatiotemporal environmental variation to quantitative genetic predictions of evolution. A difference in the phenotypic mean between ancestral and derived populations reared under common-garden conditions is strong evidence for evolution and thus is considered the “gold standard” for demonstrating evolutionary change in natural populations (Pemberton 2010; Walsh and Lynch 2018). Although by themselves common-garden experiments do not demonstrate the causes of evolutionary change (they do not distinguish changes caused by selection from those caused by drift or gene flow), they are in principle the standard against which model-based predictions of evolutionary change and field observations of changes in mean phenotype can be calibrated.

Here, we explore the problem of accurately predicting evolution under changing environmental conditions. We used animal models based on 3 years of mark-recapture data for a pedigreed, wild experimental population of Trinidadian guppies (*Poecilia reticulata*) to generate predictions of evolutionary change based on the breeder’s equation, the secondary theorem of selection, and the change in estimated breeding values. We then compared these predictions to observations of phenotypic change made in common-garden experiments as well as to the phenotypic change observed in the field data over the course of the study. The match was, in general, poor. However, when we accounted for environmental change when estimating quantitative genetic parameters, predictions based on the secondary theorem of selection and the change in estimated breeding values better matched the observation of change derived from the common-garden experiment. We discuss our results in the context of the challenges imposed by systematic environmental change when quantifying evolution.

## Methods

### *Overview of Experimental Design*

Our two objectives in this study were to assess (1) how well quantitative genetic predictions match observations of evolutionary change in a wild population and (2) whether explicitly modeling subgroup effects improves the accuracy of predictions. Our study population is an experimentally introduced wild population of Trinidadian guppies. To generate predictions, we fit animal models with mark-recapture data from our study population, then parameterized quantitative genetic models of evolutionary change.

We fit animal models both with and without subgroup effects as covariates. To observe evolutionary change, we performed annual common-garden experiments comparing traits between the introduced and ancestral populations. If quantitative genetic parameters can be accurately estimated, predictions of evolution should match the changes observed in common-garden experiments where environmental effects are controlled.

We used three quantitative genetic approaches to predict evolutionary change over a 3-year period: the univariate breeder’s equation, the secondary theorem of selection, and the change in estimated breeding values. We considered a single trait—male size at maturity—and used relative lifetime reproductive success as our measure of fitness. Using bivariate animal models, we estimated individual breeding values and additive genetic and phenotypic variances and covariances among size at maturity and fitness. We fit animal models both with and without subgroup-specific environmental factors as covariates, to determine how changes in environmental conditions over the course of the study influenced the expression of traits and the estimation of quantitative genetic parameters and breeding values. To describe the temporal dynamics of evolutionary change, we used piecewise linear regression of the change in estimated breeding values with time, allowing different slopes in each year. Finally, we calculated the expected change in trait values according to our three quantitative genetic approaches and compared these predictions to the observed change in trait values seen in (i) the study population field data (i.e., without correction of subgroup effects) and (ii) guppies reared under common-garden conditions (i.e., corrected for subgroup effects). All analyses were conducted in R (R Core Team 2018). Data and code underlying the analyses and figures in this article have been deposited in the Dryad Digital Repository (<https://doi.org/10.5061/dryad.jm63xsj7k>; Potter et al. 2020).

### *Study Population*

Trinidadian guppies are small live-bearing fish native to freshwater streams in Trinidad. Guppies display two distinct ecotypes that are defined by the intensity of local predation pressure. These ecotypes differ in a number of morphological, life-history, behavioral, and physiological traits. For example, the low-predation ecotype matures at larger sizes; has fewer, larger offspring per litter; and has lower reproductive rates than the high-predation ecotype (Reznick and Endler 1982; Reznick 1982*b*; Reznick and Bryga 1987). These differences have a genetic basis and have independently evolved in several different streams throughout Trinidad (Reznick 1982*a*; Alexander et al. 2006). The evolution of the low-predation ecotype is

driven by increased population density rather than being the direct result of reduced predation pressure (Bassar et al. 2013; Reznick et al. 2019).

Our study population is located in the Lower Lalaja, a tributary to the Guanapo River in the Northern Range Mountains of Trinidad. This population is part of a long-standing introduction experiment, described in detail elsewhere (Travis et al. 2014; Reznick et al. 2019). In brief, high-predation ecotype guppies were introduced to a closed section of a previously guppy-free, low-predation-risk stream. This introduction replicates the upstream invasion of guppies, which has occurred naturally multiple times. The introduced population has since been censused monthly. Each founder and subsequent recruits received a unique identifying mark, and two scales were removed from the caudal peduncle to sample for DNA. The mean monthly probability of capturing an individual in a census (given that it is alive) is 0.9 (Reznick et al. 2019). At each census, standard length was recorded as the distance from the tip of the snout to the hypural plate, recorded to the nearest 0.1 mm; mass was recorded to the nearest milligram, and the stage of sexual maturity in males was classified as either juvenile or mature, based on the development of the intromittent organ. Unfortunately, we do not have a similar metric for characterizing maturity status in wild-caught females, so our study is limited to considering males only.

#### *Mark-Recapture and Pedigree Data*

Our study focused on the first 3 years of the experimental introduction, starting in March 2008. Because we were interested in male size at maturity, our data set was restricted to those males caught and measured while in the juvenile stage (~14 mm standard length) and then caught and measured a month later and determined to have reached maturity. We recorded size at maturity as the standard length of males when first recorded as having reached maturity after having been observed to be immature in the preceding month. Using the individual DNA samples, we constructed a microsatellite-based pedigree for the introduced population. We used this pedigree to structure the animal models and to calculate individual lifetime reproductive success. Full details on genotyping, pedigree construction, and estimation of lifetime reproductive success are provided in the supplemental PDF (available online). Our final phenotypic data set included 851 individuals, for which accurate estimates of lifetime reproductive success and size at maturity were available. We trimmed our pedigree to exclude lineages for which we had no phenotypic data. Our final pedigree contained relatedness data for 1,506 individuals (male and female) over 3 years, with a maximum depth of seven generations. We estimated

monthly population density as the number of individuals captured in a given month's census, adjusted for capture probability, divided by the total area of the stream.

#### *Common-Garden Experiments*

Juvenile guppies (~12 mm standard length) from the introduced and ancestral populations were collected at yearly intervals following the introduction, from 2009 to 2011. To account for maternal and environmental effects, guppies were reared for two generations under controlled environmental conditions. Fish were reared with a crossing design that sustained the genetic diversity of the founders and avoided inadvertent selection by equally representing each line in each generation (Torres-Dowdall et al. 2012; Handelsman et al. 2013; Ruell et al. 2013; Reznick et al. 2019). In the common-garden experiments, guppies were reared at either high or low food levels. Because there was no effect of any interaction between food level and population (ancestral or introduced) on size at maturity (Reznick et al. 2019), we pooled data across food treatments to increase sample sizes. Male size at maturity was recorded for second-generation laboratory-born guppies.

#### *Observations of Phenotypic Change*

We had two sources for observations of phenotypic change: the common-garden data allowed us to quantify change between ancestral and introduced populations when subgroup effects were controlled, and the mark-recapture data (which were used to fit the animal models) provided measures of the phenotype in the wild, without any correction for subgroup effects.

To calculate observations of change in the common-garden data, we tested for phenotypic divergence between ancestral and derived populations with Bayesian generalized linear mixed effects models using MCMCglmm (Hadfield 2010). Full details of the model-fitting process are given in the supplemental PDF. We analyzed each year separately, including population (ancestral or derived) as a fixed effect. We accounted for relatedness among individuals within populations by fitting full-sibling family identity as a random effect. Sample sizes from the ancestral ( $N_A$ ) and introduced ( $N_I$ ) populations in each year were as follows: for 2009,  $N_A = 68$  and  $N_I = 45$ ; for 2010,  $N_A = 52$  and  $N_I = 29$ ; and for 2011,  $N_A = 41$  and  $N_I = 40$ . The difference in the mean trait in the introduced population, relative to the ancestral, is given by the coefficient for "population."

For the mark-recapture data, we used MCMCglmm (Hadfield 2010) to fit a piecewise linear regression of size at maturity against time (months since introduction), allowing different slopes in each year of the study. The

intercept of the model provides the mean initial size at maturity, and estimates of mean size at maturity can be obtained for any month during the study.

#### *Estimating Quantitative Genetic Parameters*

Using the field data ( $N = 851$ ) and pedigree, we fit bivariate animal models in the form

$$[z, w] \sim \mathbf{bX} + \mathbf{D}_1\mathbf{a} + \mathbf{D}_2\mathbf{m} + \mathbf{Ie}, \quad (3)$$

in which size at maturity ( $z$ ) and lifetime reproductive success (fitness;  $w$ ) are estimated as functions of a matrix of fixed effects ( $\mathbf{b}$ ) and random effects of breeding value ( $\mathbf{a}$ ) and maternal identity ( $\mathbf{m}$ ), which account for variation due to additive genetic and maternal effects, respectively. Breeding values were estimated on the basis of a covariance matrix structured by pairwise relatedness of all individuals in the pedigree (Lynch and Walsh 1998). Maternal identity was included to account for phenotypic similarity among siblings arising from their shared maternal environment. The terms  $\mathbf{X}$ ,  $\mathbf{D}_1$ , and  $\mathbf{D}_2$  are design matrices linking trait and fitness observations to the fixed and random effects. The residual environmental variation in traits and fitness is captured in  $\mathbf{e}$ , which is linked to individual observations by the incidence matrix ( $\mathbf{I}$ ). Animal models were fit using MCMCglmm (Hadfield 2010), generating 8,000 samples of the posterior distribution. Full details and code for the model-fitting process are given in the supplemental PDF.

By fitting bivariate models, we estimated the variance-covariance matrices among size at maturity and relative fitness due to additive genetic ( $\mathbf{G}$ ), maternal environment ( $\mathbf{M}$ ), and residual environmental ( $\mathbf{E}$ ) effects. For example, the  $\mathbf{G}$  matrix is given by

$$\mathbf{G} = \begin{bmatrix} V_A(z) & \text{cov}_A(z, w) \\ \text{cov}_A(w, z) & V_A(w) \end{bmatrix}, \quad (4)$$

where  $V_A$  and  $\text{cov}_A$  are the additive genetic variances and covariances, respectively, of  $z$  and  $w$ . The same structure is used for  $\mathbf{M}$  and  $\mathbf{E}$ , and the total phenotypic variance-covariance matrix ( $\mathbf{P}$ ) is given by  $\mathbf{P} = \mathbf{G} + \mathbf{M} + \mathbf{E}$ . This bivariate decomposition allowed us to estimate breeding values and parameters for the breeder's equation and the secondary theorem of selection in the same model. To estimate the per-generation response predicted by the breeder's equation (eq. [1]), we multiplied  $V_A(z)$  by the selection gradient, which we calculated as the phenotypic covariance between the trait and fitness divided by the total phenotypic variance, that is,

$$\beta_z = \frac{\text{cov}_A(z, w) + \text{cov}_M(z, w) + \text{cov}_E(z, w)}{V_A(z) + V_M(z) + V_E(z)}.$$

Note that because we used a log-link function to transform absolute lifetime reproductive success, estimates on

the latent scale are converted from absolute to relative lifetime reproductive success (Smouse et al. 1999; Firth et al. 2015; Bonnet and Postma 2018; Bonnet et al. 2019), meaning that our estimates of  $\beta_z$  are equivalent to selection gradients, as described by Lande and Arnold (1983). For the secondary theorem of selection (eq. [2]), the per-generation response is given by  $\text{cov}_A(z, w)$ .

#### *Temporal Dynamics in Breeding Values and Drift*

To assess the temporal dynamics of the evolution of size at maturity, we modeled the change in estimated breeding values over time. Using the estimated breeding values from our animal models, we fit piecewise linear regressions of estimated breeding values by monthly cohort. We chose this approach to allow the direction and rate of genetic change (i.e., the slopes of the model) to differ in each year of the study because our a priori expectation is of a delay in the onset of evolution of size at maturity (Reznick et al. 2019). We fit these models across all Bayesian posterior samples to account for uncertainty in estimated breeding values and the correlational structure of error terms (Hadfield et al. 2010), generating a posterior distribution of the monthly change in mean breeding value for each year (i.e., we used a piecewise linear regression modification of the approach used in, e.g., Pigeon et al. 2016; Bonnet et al. 2019). To illustrate nonlinear temporal dynamics of estimated breeding values, we also fit a general additive model with a smoothing function to each iteration of the Monte Carlo Markov chain used to sample from the posterior, as described above, using the R package mgcv (Wood 2017).

A change in breeding values may arise from genetic drift. We simulated genetic drift by generating 8,000 sets of random breeding values, based on the posterior distribution of  $V_A(z)$  and the pedigree (Hadfield et al. 2010). We then fit the piecewise linear regression model described above using these random breeding values, to simulate the change in estimated breeding values that could be explained by genetic drift. We calculated the posterior probability that changes in mean estimated breeding values ( $\Delta\bar{eBV}$ ) were greater than what would be expected as a result of genetic drift ( $\Delta\bar{eBV}_d$ ) as the proportion of posterior samples of  $\Delta\bar{eBV}$  that were greater than the equivalent posterior samples of  $\Delta\bar{eBV}_d$  (e.g., as in Hadfield et al. 2010).

#### *Including Subgroup Effects as Covariates*

Because we wanted to test how subgroup effects can influence estimation of quantitative genetic parameters, we ran two types of model. In the first, the only fixed effects were the population mean values for each response variable. Our predictions for the noncorrected breeder's equation, secondary theorem of selection, and change in

estimated breeding values were derived from this model. In the second, we also included subgroup-specific measures of environmental variables as fixed effects. This allowed us to test how the expression of traits and fitness are influenced by known environmental variables and to determine how correcting phenotypes for subgroup effects influenced estimation of breeding values and quantitative genetic parameters.

We defined a subgroup of guppies as those that attained maturity in the same month, allowing us to account for broad-scale temporal changes in environmental factors influencing the population. Our subgroup-specific environmental factors were population density, season (rainy or dry), and their interaction, which we included as fixed effects in the animal model. We expect that population density and season affect resource availability, which in turn is likely to influence the means and variances of traits and fitness. Trinidad has a pronounced rainy season from June to December, during which flooding scours the streams of invertebrates and algae, reducing the availability of food for guppies. Population density is also likely to affect resource density, but we expect the relationship between population density and resource availability to change as the carrying capacity changes between seasons.

A potential issue with correcting our measure of fitness for population density is that population density is a function of mean fitness: over a single time step, population growth rate and mean absolute fitness are equivalent (Engen et al. 2009). However, our measure of population density applies to the month in which male guppies attained maturity, which is by definition before those individuals have contributed to population growth with their own offspring. This decouples individual lifetime reproductive success from the population density measure applied to that individual.

We tested whether subgroup effects differed between generations by calculating the mean difference in population density experienced at maturity between sires and their offspring. We did this by fitting a linear model of population density with relation (offspring or sire) as a fixed effect, and we included sire-offspring pair as a grouping factor (i.e., random effect). The intercept of the model is the mean population density experienced by offspring, and the slope gives the mean difference in population density experienced by sires. A negative value of the slope indicates that sires experienced lower population densities than their offspring. We fit this model using MCMCglmm (Hadfield 2010), with default settings and priors.

Covariance between a trait and fitness may arise as a result of (a) a causal effect of the trait on fitness, (b) selection on a correlated (unmeasured) trait, and/or (c) environmental covariance between the trait and fitness (Walsh and Lynch 2018). When the covariance is generated by

*b* and/or *c*, the estimate of the causal effect of the trait on fitness (i.e., the phenotypic selection gradient) will be biased. Bias can be detected by decomposition of the phenotypic selection gradient  $\beta_z$  into additive genetic ( $\beta_A = \text{cov}_A(z, w)/V_A(z)$ ) and environmental ( $\beta_E = \text{cov}_E(z, w)/V_E(z)$ ) selection gradients; selection is biased when  $\beta_A \neq \beta_E$  (Rausher 1992; Stinchcombe et al. 2002; Hadfield 2008). If the focal trait is the sole determinant of variation in fitness (i.e., the true target of selection), then  $\beta_z = \beta_A$ , and predictions from the breeder's equation and the secondary theorem of selection will be equivalent (Morrissey et al. 2010, 2012; Walsh and Lynch 2018). We assessed the bias in our estimates of selection by calculating the posterior probabilities that  $\beta_E > \beta_A$  and  $\beta_z > \beta_A$ . By fitting an animal model with subgroup effects as covariates, we estimated  $\beta_z$  conditional on the subgroup effects, thus removing the effect of *c* due to changes in population density and season. Remaining bias would indicate that unmeasured correlated traits (or other environmental factors) also influence variation in fitness.

#### Comparing Predictions and Observations

We assume that the difference in trait values between ancestral and introduced guppies in the common-garden experiments reflects the true genetic change. To compare predictions of change with observations of change in the field and the common-garden experiments, we multiplied the per-generation predictions from the breeder's equation and the secondary theorem of selection by the mean number of generations elapsed after 3 years (mean generation time was 7.90 months;  $N = 1,457$ ,  $SD = 3.15$ ; see the supplemental PDF for details on this calculation). Because the common-garden data and the field data used to fit the animal models are independent, we could directly calculate the posterior distribution of the differences between predictions and observations. For example, the difference between prediction A and observation B was obtained by subtracting the posterior distribution of B from that of A. The resultant distribution of the difference,  $|A - B|$ , captures the magnitude and uncertainty of the difference between estimates. Note that due to a slight skew in posterior distributions, the mode of  $|A - B|$  may not be equal to the mode of A minus the mode of B. We define a difference (e.g.,  $|A - B|$ ) as statistically clear when the 95% credible intervals of the distribution do not include zero.

## Results

### Observations of Phenotypic and Environmental Change

Parameter estimates provided in the following sections are the modes of the posterior distributions of estimates, with 95% credible intervals given in parentheses or brackets.

Male size at maturity as measured in the field was highly variable (fig. 1A). At the onset of the experiment, mean size at maturity ( $\bar{z}$ ) was 17.66 mm (17.17 to 18.07);  $\bar{z}$  declined after 2 years (after 1 year:  $\bar{z} = 17.62$  mm [17.47 to 17.78]; after 2 years:  $\bar{z} = 16.29$  mm [16.18 to 16.43]), before increasing to slightly less than the initial size in the third year (after 3 years:  $\bar{z} = 17.41$  mm [17.21 to 17.61]). Mean total population density increased each year following the introduction (fig. 1B; year 1: mean =  $0.56 N m^{-2}$ , SD =  $0.25 N m^{-2}$ ,  $N = 11$ ; year 2: mean =  $1.69 N m^{-2}$ , SD =  $0.38 N m^{-2}$ ,  $N = 12$ ; year 3: mean =  $2.55 N m^{-2}$ , SD =  $0.36 N m^{-2}$ ,  $N = 12$ ).

In the common-garden experiments, mean size at maturity relative to the ancestral population increased slightly in the first year of the study, then decreased slightly in the second year, followed by a clear increase in the third year (i.e., 95% credible intervals of the change do not overlap zero; fig. 2). By contrast, mean size at maturity as measured in the field was clearly lower only in the second year of the experiment (fig. 2). Observations of change in size at maturity clearly differed between the field and the common garden in years 2 and 3 ( $|\Delta z_{CG2} - \Delta z_{F2}| = 1.00$  mm [0.36 to 1.70];  $|\Delta z_{CG3} - \Delta z_{F3}| = 0.70$  mm [0.07 to 1.31]) but not in year 1 ( $|\Delta z_{CG1} - \Delta z_{F1}| = 0.34$  mm [−0.33 to 1.03]; fig. 2).

#### *Environmental Influence on Size at Maturity and Fitness*

The environmental subgroup effects that we measured—population density and season—strongly influenced the expression of both size at maturity and fitness. The size at which male guppies attained maturity decreased as population density increased ( $\beta_{pop} = -0.66$  mm per  $N m^{-2}$  [−0.80 to −0.51]). Guppies tended to mature at smaller sizes during the rainy season ( $\beta_{rainy} = -0.72$  mm [−1.09 to −0.37]). There was an interaction between population density and season, such that the negative effect of increasing population density on size at maturity was around 1.7 times greater during the dry season than during the rainy season ( $\beta_{pop,rainy} = 0.33$  mm per  $N m^{-2}$  [0.15 to 0.51]).

The relative lifetime reproductive success of male guppies was influenced by the environmental conditions they experienced in the month prior to maturity. Guppies had lower fitness if they attained maturity at times of increased population density ( $\beta_{pop} = -0.59$  per  $N m^{-2}$  [−0.85 to −0.33]) and tended to have lower fitness if they attained maturity during the rainy season ( $\beta_{rainy} = -0.67$  [−1.30 to −0.03]). Again, there was an interaction between these factors, meaning that the decrease in lifetime reproductive success with increasing population density was strongest during the dry season ( $\beta_{pop,rainy} = 0.33$  per  $N m^{-2}$  [0.01 to 0.65]).

Subgroup effects differed across generations: on average, the population densities experienced by offspring were ~40% greater than those experienced by their sires (mean difference in population density at maturity of sire relative to that of offspring =  $-0.59 N m^{-2}$  [−0.63 to −0.55]; fig. S3; figs. S1–S3 are available online).

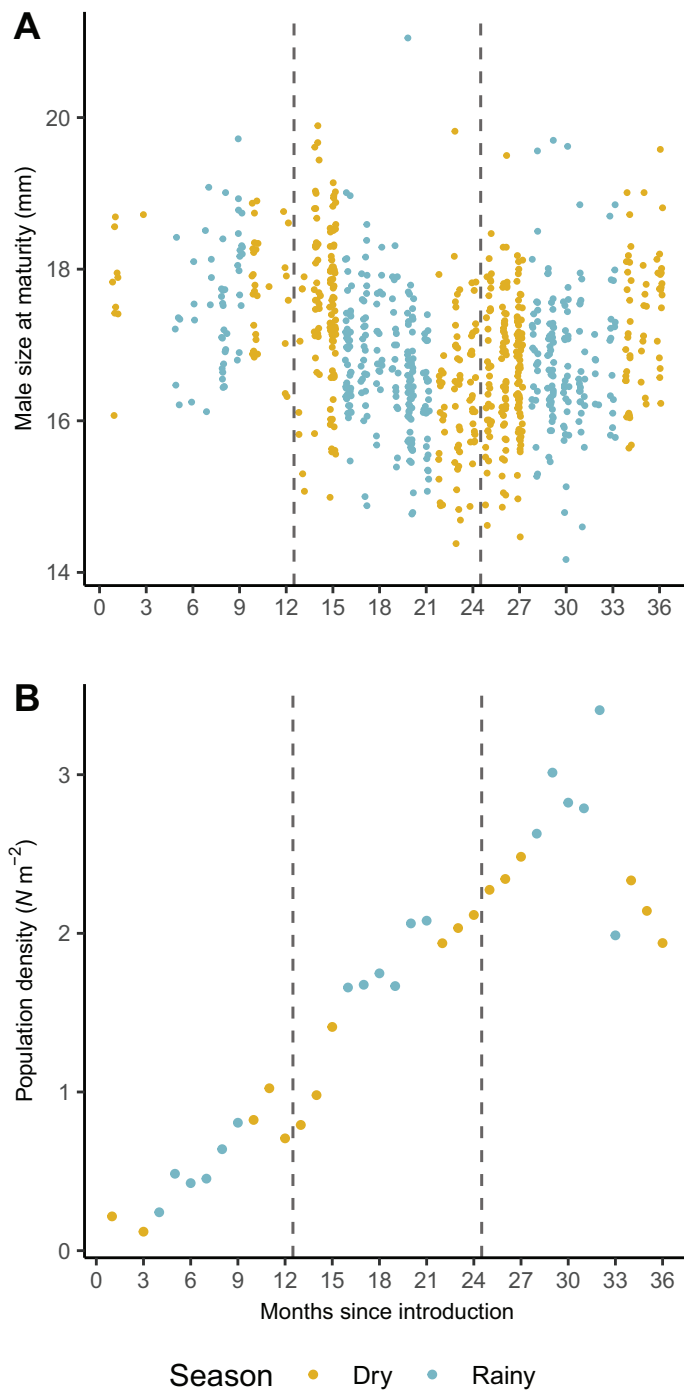
#### *The Impact of Subgroup Effects on Estimation of Quantitative Genetic Parameters*

We ran two animal models that differed in whether subgroup effects were included as covariates. In both models, we detected significant additive genetic variance ( $V_A$ ) for size at maturity  $z$  and fitness  $w$  (table 1). Maternal effects accounted for small but significant proportions of the total phenotypic variance ( $V_P$ ) in both  $z$  and  $w$  (table 1). Residual environmental variance explained the majority of  $V_P$  in both size at maturity and relative fitness (table 1).

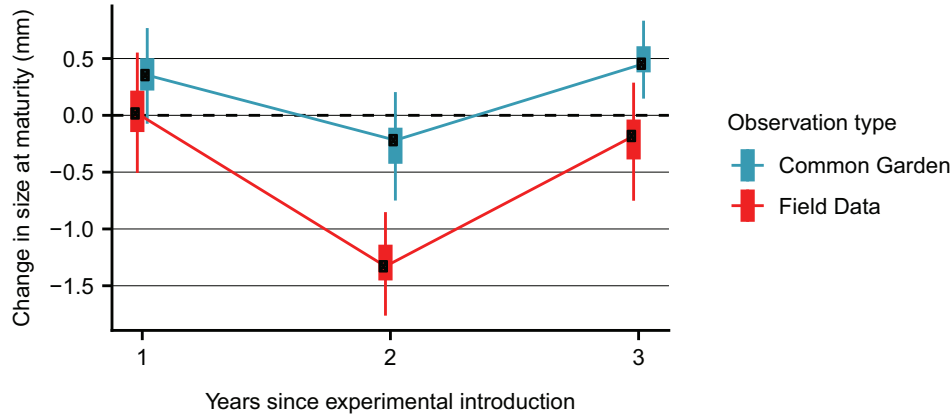
If relatives experience different subgroup effects, estimation of additive genetic (co)variances and breeding values may be biased when subgroup effects are not modeled (i.e., the first problem outlined in the introduction). We found little difference between estimates of  $V_A(z)$  between models fit with and without subgroup effects, but  $V_A(w)$  was around 50% greater when subgroup effects were included in the animal model (table 1). Accounting for subgroup effects had a pronounced effect on the estimation of the additive genetic covariance between size at maturity and fitness  $cov_A(z, w)$ : when subgroup effects were included, the sign of  $cov_A(z, w)$  changed from negative to positive (table 1). Accordingly, when subgroup effects were not modeled, the secondary theorem of selection predicted a per-generation decrease in mean size at maturity of 0.09 mm (−0.07 to 0.25), but it predicted an increase of 0.11 mm (−0.07 to 0.27) per generation when subgroup effects were accounted for. Note that while the credible intervals for both of these estimates span zero, the former indicates an 85.2% posterior probability that  $z$  will decrease, while the latter indicates an 86.6% posterior probability that  $z$  will increase.

When unmodeled subgroup effects vary over time, changes in estimated breeding values may reflect phenotypic rather than genetic change. For both models, the change in estimated breeding values over the duration of the study was nonlinear but could be approximated as linear on an annual basis (fig. 3). Without modeling subgroup effects, the change in estimated breeding values resembled the phenotypic change seen in the field data: there was no clear change in mean breeding value in the first year ( $\Delta \bar{BV}_1 = 0.01$  mm month<sup>−1</sup> [0.00 to 0.02]), a decrease in the second year ( $\Delta \bar{BV}_2 = -0.03$  mm month<sup>−1</sup> [−0.05 to −0.01]), and an increase in the third year ( $\Delta \bar{BV}_3 = 0.05$  mm month<sup>−1</sup> [0.02 to 0.07]; fig. 3A;





**Figure 1:** Raw data showing changes in male size at maturity (A) and population density over time (B). Data are shown on a monthly basis for the first 3 years of the introduction experiment. Data points in A are jittered on the X-axis to facilitate visualization of similar values. Vertical dashed lines highlight changes in year of study. Observations in the dry season are given in yellow, and observations in the rainy season are given in blue.



**Figure 2:** Annual observations of mean change size at maturity in the common-garden experiments (blue) and in the mark-recapture field data (red). Changes in the common garden are based on comparisons between the grandoffspring of individuals sampled from the ancestral and experimental populations each year, reared under common conditions. Observations of change in the field data are of size at maturity, as measured in the experimental population, and fit with a piecewise linear model, allowing for different slopes in each year.

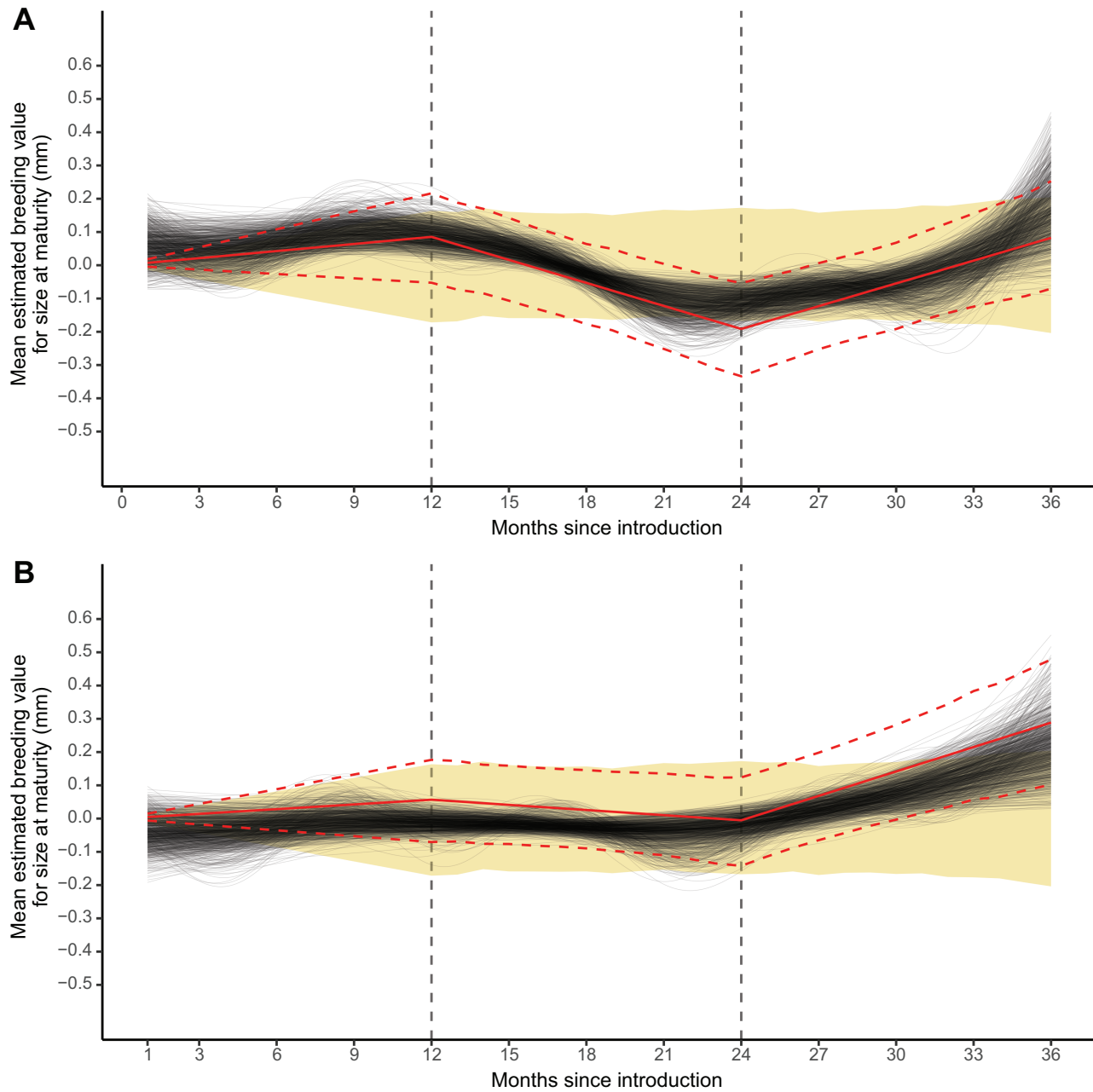
cf. fig. 2). By contrast, when we modeled subgroup effects, the pattern in the breeding values resembled that in the common-garden experiments: there was no clear change in the first 2 years ( $\Delta\overline{BV}_1 = 0.00 \text{ mm month}^{-1} [-0.01 \text{ to } 0.01]$ ;  $\Delta\overline{BV}_2 = -0.01 \text{ mm month}^{-1} [-0.03 \text{ to } 0.01]$ ), but there was a clear increase in the third year ( $\Delta\overline{BV}_3 = 0.03 \text{ mm month}^{-1} [0.01 \text{ to } 0.05]$ ; fig. 3B; cf. fig. 2). After 3 years, the posterior probability that the increase in estimated breeding values was greater than what would be expected as a result of genetic drift was 98.0%.

If subgroup effects are present but not modeled, the phenotypic selection gradient may capture a noncausal association between the trait and fitness (i.e., the second problem outlined in the introduction). Without modeling subgroup effects, we found that male guppies that matured at a larger size had higher fitness, indicating positive phenotypic selection for this trait ( $\beta_z = 0.16 \text{ mm}^{-1} [0.05 \text{ to } 0.28]$ ; fig. S2). For comparison with other studies, the standardized selection gradient (Kingsolver et al. 2001) was 0.16 (0.05 to 0.29); the mean-standardized selection

**Table 1:** Components of variances and covariance from animal models fit with and without subgroup effects as covariates

Trait, parameter	Estimate	
	Without subgroup effects	With subgroup effects
Size at maturity ( $z$ ):		
$V_A$	<b>.36</b> (.21 to .56)	<b>.32</b> (.17 to .47)
$V_M$	<b>.10</b> (.05 to .20)	<b>.09</b> (.04 to .16)
$V_E$	<b>.51</b> (.39 to .66)	<b>.46</b> (.35 to .60)
$V_z$	<b>1.01</b> (.91 to 1.13)	<b>.89</b> (.81 to .99)
Relative fitness ( $w$ ):		
$V_A$	<b>.47</b> (.18 to .84)	<b>.73</b> (.36 to 1.31)
$V_M$	<b>.12</b> (.04 to .29)	<b>.10</b> (.04 to .27)
$V_E$	<b>1.24</b> (.90 to 1.63)	<b>.97</b> (.60 to 1.38)
$V_z$	<b>1.89</b> (1.61 to 2.28)	<b>1.92</b> (1.62 to 2.29)
cov( $z, w$ ):		
cov $_A(z, w)$	-.09 (-.25 to .08)	.11 (-.07 to .28)
cov $_E(z, w)$	<b>.20</b> (.05 to .35)	-.02 (-.18 to .13)
cov $_z(z, w)$	<b>.17</b> (.04 to .28)	.11 (-.01 to .22)

Note: Posterior modes and 95% credible intervals (in parentheses) are provided for estimates of variances ( $V$ ) and covariances (cov) relating to size at maturity ( $z$ ) and relative fitness ( $w$ ). We used lifetime reproductive success as our measure of fitness. Subscripts A, M, and E denote, respectively, the additive genetic, maternal environment, and residual environmental components of the total phenotypic variance  $P$ . Note that calculations of  $V_P$  and cov $_z(z, w)$  were performed on the posterior distributions of their components; the sum of the modes of components does not necessarily equal the mode of  $V_P$ . Estimates for which 95% credible intervals do not overlap zero are given in boldface.



**Figure 3:** Change in mean estimated breeding value for size at maturity over the 36-month study period without correcting for environmental change (A) and when changes in population density and season are accounted for when estimating breeding values (B). Solid red lines are the mean trend in estimated breeding values and dashed red lines indicate the 95% credible intervals of the trend, based on piecewise linear regression of the change in breeding value over time. Each gray line is the mean estimated breeding value fit with a general additive model to a single iteration of the Monte Carlo Markov chain; 1,000 lines are shown to illustrate the uncertainty around the non-linear trend in estimated breeding values with time. The yellow region is the 95% credible interval of the change in breeding values expected as a result of drift. Vertical dashed lines highlight changes in year of study.

gradient (Hereford et al. 2004) was 2.69 (0.77 to 4.93). However, decomposition of  $\beta_z$  revealed that the genetic ( $\beta_A$ ) and environmental ( $\beta_E$ ) selection gradients were not equal, meaning that this estimate of selection was biased

( $\beta_A = -0.22 \text{ mm}^{-1} [-0.70 \text{ to } 0.22]$ ;  $\beta_E = 0.37 \text{ mm}^{-1} [0.07 \text{ to } 0.70]$ ;  $\text{Pr}(\beta_A < \beta_E) = 95.7\%$ ). After correcting for subgroup effects, the strength of phenotypic selection was slightly reduced ( $\beta_{z|SG} = 0.12 \text{ mm}^{-1} [-0.01 \text{ to}$

0.25]); although the confidence interval includes zero, there was a 97.0% posterior probability that selection was positive. This corrected estimate  $\beta_{z|SG}$  accounted for any potential noncausal covariance generated by subgroup effects but was still biased. Phenotypic and genetic selection gradients were not equal ( $\beta_{A|SG} = 0.30 \text{ mm}^{-1} [-0.24 \text{ to } 0.89]$ ;  $\Pr(\beta_{A|SG} > \beta_{z|SG}) = 77.7\%$ ), implying that unmeasured genetically correlated traits (or other unmodeled environmental factors) caused variation in fitness.

Because there was little difference between models in our estimates of phenotypic selection and additive genetic variance, predictions of evolution according to the breeder's equation were similar whether subgroup effects were accounted for or not. Without accounting for subgroup effects, the breeder's equation predicted an increase of 0.06 mm per generation (0.01 to 0.11); when subgroup effects were modeled, an increase of 0.04 mm per generation was expected (0.00 to 0.09).

#### *Comparing Predictions with Observations from the Common Garden*

When we did not include subgroup effects as covariates in the animal model, all three predictions of change underestimated the change seen in the common-garden experiments after 3 years (fig. 4, in red). The predicted change according to the secondary theorem of selection was in the opposite direction to the change seen in the common-garden experiments and so underestimated the observed change by 0.93 mm (0.10 to 1.73); the estimated breeding value approach underestimated the change by 0.40 mm (0.04 to 0.78), and the breeder's equation underestimated the change by 0.22 mm ( $-0.20$  to 0.62), although this last difference was not statistically clear.

When we accounted for subgroup effects, there were no statistically clear differences between predictions and observations in the common garden ( $|\Delta\bar{z}_{\text{pred}} - \Delta\bar{z}_{\text{obs}}|$ ; fig. 4, in blue). For the secondary theorem of selection,  $|\Delta\bar{z}_{\text{STS}} - \Delta\bar{z}_{\text{obs}}|$  was reduced to  $-0.02$  mm ( $-0.93$  to 0.83); for the estimated breeding value approach,  $|\Delta\bar{z}_{\text{EBV}} - \Delta\bar{z}_{\text{obs}}|$  was reduced to  $-0.19$  mm ( $-0.59$  to 0.20). Correcting for subgroup effects decreased the accuracy of the breeder's equation prediction:  $\Delta\bar{z}_{\text{BE}} - \Delta\bar{z}_{\text{obs}}$  increased to  $-0.34$  mm ( $-0.73$  to 0.07).

### Discussion

#### *Accounting for Environmental Effects Alters Predictions about Evolution*

Without accounting for subgroup effects, our results appeared to be a classic case of evolutionary stasis: we identified additive genetic variance and positive selection for size at maturity but actually observed a slight decline in

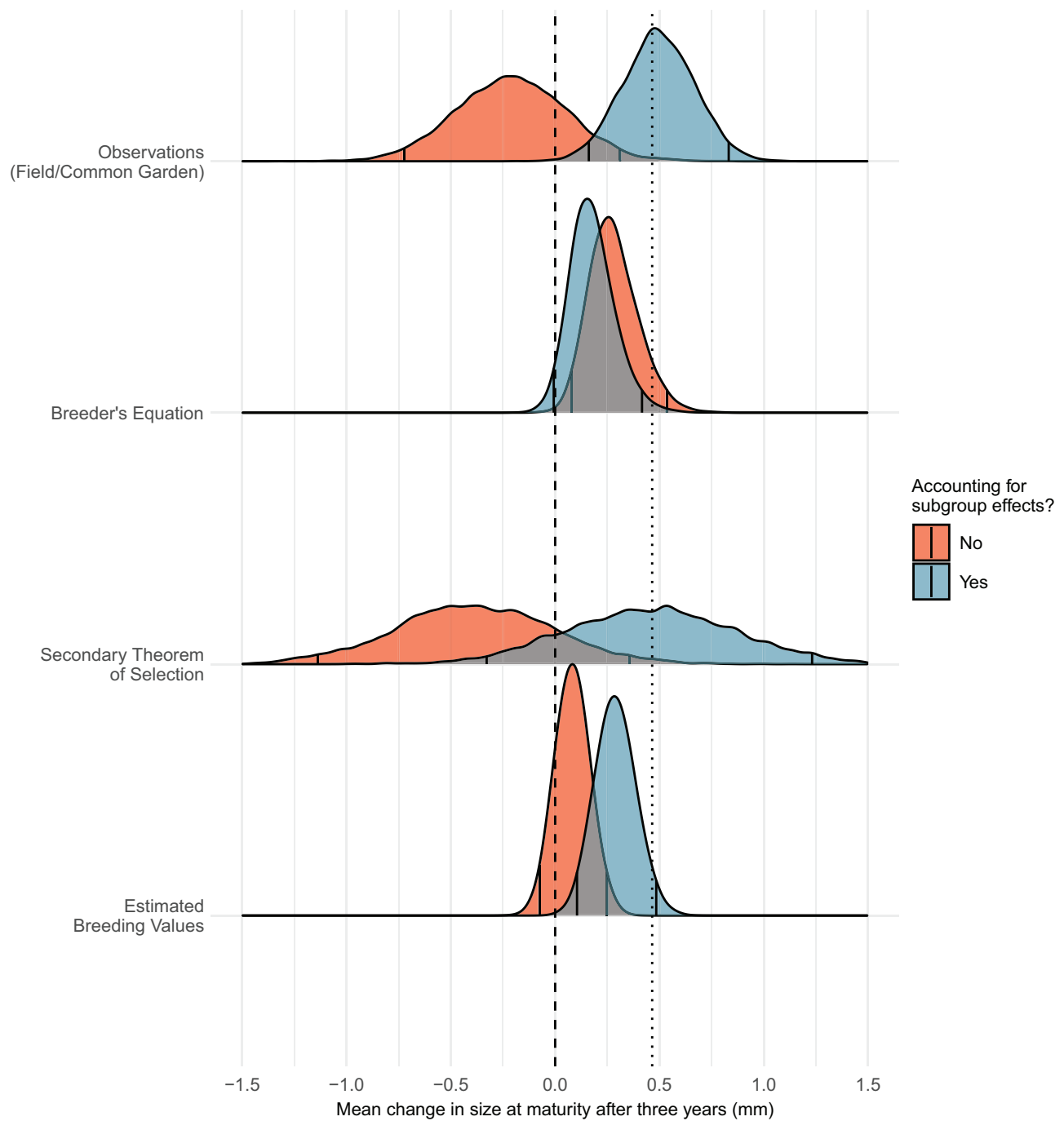
this trait in the field phenotypic data (fig. 4, in red). Using the secondary theorem of selection, we then detected slightly negative selection occurring at the genetic level (albeit with considerable uncertainty), which agreed with the limited change seen in the field data (fig. 4, in red). This type of discrepancy between predictions of change using the breeder's equation and the secondary theorem of selection has previously been attributed to unmeasured factors upwardly biasing estimates of selection at the phenotypic level; that is, if traits that influence fitness are not measured, the breeder's equation may predict evolution when none has occurred (Morrissey et al. 2012). Initial analysis of the estimated breeding values also indicated the absence of a clear evolutionary change after 3 years (fig. 4, in red).

However, results from the common-garden experiments show that mean size at maturity in the experimental population increased relative to the ancestral population: this increase was roughly double that predicted by the breeder's equation (fig. 4). Given that the natural low-predation guppy ecotype matures at larger sizes than the high-predation ecotype (Reznick and Endler 1982), this change is consistent with the previously observed evolutionary transitions between ecotypes. Because environmental effects were controlled in the common garden, this increase must be due to an increase in the mean breeding value of the population, that is, evolution.

This result begs the question: why did these quantitative genetic models underestimate or fail to predict the evolutionary change demonstrated by the common-garden experiments? Our results suggest that not accounting for environmental change over the study period (i.e., subgroup effects) led us to predict an incorrect evolutionary trajectory. When we modeled subgroup effects (i.e., differences among subgroups in the population density and season experienced prior to maturity), predictions of change in size at maturity based on the secondary theory of selection and the change in breeding values better matched observations in the common-garden experiments, although considerable uncertainty remained (fig. 4, in blue).

How then did changes in population density and season bias our estimation of quantitative genetic parameters? In the introduction we outlined two distinct problems when such subgroup effects are not modeled: (i) environmentally induced dissimilarity among relatives and (ii) noncausal associations between traits and fitness. In the following paragraphs, we discuss our results in the context of these problems.

Regarding the first problem, when we did not model subgroup effects, we implicitly assumed that relatives experienced the same environmental conditions. This was not the case: because our study population was growing, offspring experienced considerably higher population



**Figure 4:** Observed and predicted change in mean size at maturity in male guppies, 3 years after their experimental introduction to a low-predation-risk environment. Observations and predictions that do not account for subgroup effects are given in red; those that do are given in blue. Observations are of phenotypic change in the mark-recapture field data (red) and in common-garden experiments (blue). The observed change in the field data is the mean change observed after 36 months, based on a piecewise linear regression of phenotypic change over time. The observed change in the common-garden data is the mean difference between the grandoffspring of individuals sampled from the ancestral and experimental population 3 years after the introduction and reared under common conditions. Predictions of change based on the breeder's equation and the secondary theorem of selection are based on 36 months of data and assume a mean generation time of 7.90 months. Predictions based on the change in estimated breeding values are derived from piecewise linear regression of estimated breeding values over time. For predictions, subgroup effects were accounted for by including population density, season, and their interaction as fixed effects in the animal model used to estimate quantitative genetic parameters and breeding values. Full posterior distributions of observations and predictions of phenotypic change are shown; vertical black lines indicate the 95% credible intervals of the distributions. The vertical dotted line highlights the mean change in breeding value as determined by the common-garden experiments.

densities than their sires. At high population densities, guppies matured at smaller sizes and had lower fitness. By estimating the additive genetic covariance matrix across the full period of the study—without modeling subgroup effects—we failed to account for environmentally induced dissimilarity in phenotypes and fitness among relatives. Although this had little effect on estimates of  $V_A(z)$ , there was a change in the sign of  $\text{cov}_A(z, w)$  and a change in the temporal pattern of estimated breeding values when subgroup effects were modeled. In our first animal model, the secondary theorem of selection response was biased toward the phenotypic trend: after 3 years, the prediction of a slight decrease in mean size at maturity matched what we observed in the field measurements of this trait (fig. 4). The change in estimated breeding values also resembled the phenotypic change over time observed in the field data (fig. 3A; cf. fig. 1A). However, when we corrected traits and fitness for subgroup effects, predictions of evolution from both of these approaches more closely matched the changes seen in the common-garden experiments, that is, the genetic change (figs. 3B, 4).

Regarding the second problem, accounting for subgroup effects had little effect on estimates of the phenotypic selection gradient, but the remaining mismatch between the genetic and environmental components of the gradient indicated that correlated traits that influence fitness were missing from the analysis. However, by correcting for subgroup effects, we showed that size at maturity had a (partial) causal effect on fitness. The omission of genetically correlated traits influencing fitness from the analyses may explain the underestimation of evolutionary change predicted by the breeder's equation from both animal models (fig. 4). This is supported by the underestimation we observed in the change in breeding values when subgroup effects were modeled (fig. 4); estimation of breeding values is also sensitive to the omission of genetically correlated traits.

Both statistical problems that we have discussed are potential examples of the Yule-Simpson effect—a change in the magnitude or sign of a covariance conditional on a third variable (Yule 1903; Simpson 1951). This is a well-known problem in regression analyses, whereby covariances estimated across aggregated data capture noncausal associations when the data are structured by a third confounding variable (Blyth 1972; Pearl 2014). With regard to estimating phenotypic selection gradients (the second problem), this issue is partially solved by the development of the multivariate breeder's equation (Lande and Arnold 1983). This solution is only partial because fitness, which is convenient to think of as a trait in quantitative genetic analyses, is really an emergent property of the interaction between the phenotype and the environment (Coulson et al. 2006). As such, fitness is conditional not only on

the traits that comprise an individual's phenotype but also on the environmental conditions it experiences.

The strength of the animal model—that it makes use of relatedness data across the entire population—also makes it prone to the Yule-Simpson effect when relatives experience different environmental conditions. In our study, Yule-Simpson effects driven by two distinct unmodeled factors—differences among relatives in environmental conditions experienced and unmeasured traits influencing fitness—reduced our ability to accurately quantify evolution. The secondary theorem of selection was sensitive to the former, the breeder's equation was sensitive to the latter, and the change in estimated breeding values was sensitive to both. Our results demonstrate that such effects, if unaccounted, can lead to incorrect inferences about both the magnitude and the direction of evolutionary change.

#### *Ecological Feedback Drives Cryptic Evolution*

Our results support the hypothesis that the life histories of guppies evolve in response to increased population density rather than as a direct response to reduced predation risk (Reznick et al. 2019). If selection for increased size at maturity was due to release from predation pressure, selection should be strongest at the onset of the experiment (Reznick et al. 2019). Instead, several lines of evidence point toward size at maturity evolving in response to increases in population density: (i) the mean estimated breeding value for size at maturity only began to increase in the third year of the study, when population density was highest (fig. 3B); (ii) the monthly phenotypic trend shows a strong decline in size at maturity with increasing population density in the second year but an increase in size at maturity in the third year despite higher average population densities than in the preceding years (fig. 1); and (iii) a clear increase in size at maturity was absent in the common-garden experiments until the third year of the study (fig. 2).

Comparing change in size at maturity between the field and common-garden data revealed cryptic evolution of this trait: an increase in the genetic component of the phenotype concealed by a decrease in the environmental component (Cooke et al. 1990). Under the additive genotype-phenotype map ( $P = G + E$ ), increasing the breeding value  $G$  will increase the trait value  $P$ , provided that the environmental component  $E$  is constant (as in the common-garden experiments). In the field data, a negative value of  $E$  masked the increase in  $G$ , such that size at maturity after 3 years did not differ from that at the onset of the experiment. The effect of the increase in  $G$  can be seen in figure 1A: in the third year, the phenotypic value began to increase, despite continued increases in

population density (and by extension an increasingly negative value of  $E$ ).

The population appeared to reach carrying capacity in the third year and then began to decline. This provided an opportunity for a comparison: population densities are approximately equal at months 23–24 and months 35–36 (fig. 1B), yet there is an increase in mean size at maturity of around 1 mm between these time points. This increase is roughly equivalent to the increase in  $G$  in the common-garden data over the same period (0.78 mm). The increase in population density led to a decrease in mean size at maturity due to changes in  $E$ , but it also drove positive selection, resulting in an increase in the breeding value  $G$  of this trait.

In biological terms, as the population approached carrying capacity, per capita resource availability presumably declined, which had two conflicting effects on the population's mean phenotype. At the genetic level, decreased resource availability drove selection for increased size at maturity. This was likely because larger size in general is associated with higher competitive ability under resource limitation in guppies (Potter et al. 2019). This genetic change was only apparent in the common-garden experiments, where resource levels were constant. At the same time, decreased per capita resource availability likely had a direct effect that decreased size at maturity. This is consistent with experimental results on resource-dependent plasticity in guppies (Reznick 1990). As such, at the phenotypic level wild guppies could not attain the larger size that their breeding values would predict because there were insufficient resources to do so. Thus, our study provides a novel example of ecological feedback driving cryptic evolution.

#### *Generality and Limitations of the Study*

Our results, like most quantitative genetic studies of natural systems, are based on a single population. This makes it difficult to assess the generality of our findings. However, our study population is one of four experimental introductions: we do know that common-garden studies indicated a similar evolutionary change driven by higher densities in the other three experimental populations (Reznick et al. 2019). Barriers to dispersal, which bound the introduced populations, limit the potential for evolution due to gene flow, and our analysis of estimated breeding values indicate that the observed change was greater than that expected as a result of genetic drift. As such, our results strongly suggest that the evolution of size at maturity is an adaptive response to resource limitation. It is possible that some component of the change seen in the common-garden experiments is due to a genotype-by-environment interaction: differences in environmen-

tal conditions between the laboratory and field may exaggerate or minimize genetic differences between ancestral and introduced guppies. However, given the absence of any interaction between food level and population in the common-garden experiments in any of the years of study, the simplest explanation is an increase in mean breeding value for size at maturity occurring between years 2 and 3 of the experiment (as in fig. 3B).

Systematic changes in resource availability generating subgroup effects are likely widespread and may emerge from concurrent ecological and evolutionary dynamics. However, our experimental introduction replicated a colonization event, and our focal population of guppies was not at equilibrium. In these first 3 years, we captured the beginning of the transition from one evolutionary optimum to another, that is, from the high-predation ecotype to the low-predation ecotype. Population density increased almost 30-fold in less than 3 years (fig. 1B). Such large changes in population size are atypical of wild pedigreed populations. As such, the temporal trends we saw may have been more pronounced than in other systems. However, where systematic environmental change occurs—for example, in studies of evolution in response to climate change—we argue that accounting for subgroup effects will improve predictions of evolutionary trajectories.

One problem with our study system is that it is not feasible to perform mark-recapture on juvenile guppies smaller than 14 mm in length. This means that we have no record of individuals that died before we had an opportunity to sample them. The survival rate of juvenile guppies to a size of 14 mm in natural low-predation-type streams is estimated at ~20% (Reznick et al. 1996), meaning that we have a large “invisible fraction” (Grafen 1988). This is further compounded by missing data from individuals that were initially sampled but died before attaining maturity. This is problematic because individuals that did not survive to maturity are not included in the analysis (Hadfield 2008). If, for example, viability selection is operating on juvenile size and juvenile size is genetically correlated with size at maturity, then we will have underestimated the strength of (indirect) selection for size at maturity. This problem is offset to some extent because we also likely underestimated the change in size at maturity in the common-garden experiments relative to that which would occur in the field. In the common-garden experiment, there was no invisible fraction lost between sampling (at ~12 mm) and maturity. This means that smaller individuals that might have been lost to viability selection in the wild survived to maturity (at presumably smaller sizes) in the common-garden experiment.

Estimation of genetic (co)variances is notoriously data hungry, and subsequently there was considerable uncertainty around some of our predictions. In particular,

predictions based on the secondary theorem of selection had very wide credible intervals spanning zero. Can we then justify saying that predictions improved when environmental effects were accounted for? We would argue that the accuracy of predictions improved because the difference between predictions and observations decreased when environmental effects were accounted for. Our predictions, as is typical of such studies, lacked precision. However, by accounting for subgroup effects, predictions based on the secondary theorem of selection and the change in estimated breeding values more closely matched observations in the common garden and less closely matched the phenotypic trend in the field data.

#### *Implications for Predicting Evolution in Changing Environments*

Unaccounted environmental effects are known to bias quantitative genetic predictions of evolution (Lande and Arnold 1983; Rausher 1992; Lynch and Walsh 1998; Postma 2006). We were able to test for bias caused by the changing environment by comparing sets of different predictions for a wild population to observations in a common-garden experiment, where environmental effects were controlled. How should researchers tackle the problem of environmental change in the absence of common-garden data? Our results show that when relatives experience distinct environmental conditions, these environmental variables should be included as fixed effects in the animal model. This reduces Yule-Simpson effects by accounting for environmentally caused differences in the expression of traits and fitness prior to estimating breeding values and the additive genetic variance-covariance matrix.

In our study system, we expected selection for the low-predation guppy ecotype to be driven by increased population densities and decreased resource availability (Bassar et al. 2013; Reznick et al. 2019). Given the many factors that may influence phenotypes and fitness in wild populations, selecting an appropriate environmental variable to include in quantitative genetic analyses may be challenging. One proposed solution to this problem is to simply account for all annual environmental differences by including year as a fixed effect in the animal model (Postma 2006; Walsh and Lynch 2018). However, this confounds environmental change with any true genetic change that occurs within a year, reducing the power to detect evolution (Postma 2006; Walsh and Lynch 2018). This is particularly problematic for organisms with short generation times. We suggest that researchers carefully consider which environmental variables might cause dissimilarity in phenotype and fitness among relatives and to include those factors as covariates when estimating quantitative genetic parameters.

A general implication of our findings is that environmental drivers of selection can inhibit our ability to detect the evolutionary change they cause. Not accounting for environmental drivers of selection in quantitative genetic models may result in false-negative results—that is, the failure to detect actual evolutionary change, as in this study. This is most likely when environmentally induced phenotypic plasticity and selection act in opposing directions. Because evolutionary change is fastest when it is cryptic (Coulson et al. 2019), failure to detect instances of cryptic evolution will result in underestimation of mean rates of evolution in wild populations. Not accounting for changing environmental influences on traits and fitness reduces the ability of quantitative genetic models to accurately detect evolutionary change.

#### **Acknowledgments**

We gratefully acknowledge the contributions of the many interns, field managers, and laboratory technicians of the Guppy Project, without whom this study would not have been possible. We thank the Ramlal family in Trinidad for providing housing and logistical support. Jennifer Lau, Jill T. Anderson, Jarrod Hadfield, and an anonymous reviewer provided insightful feedback that greatly improved the manuscript. Thanks to Anja Felmy for valuable discussions and comments on this work. T.P. is funded by the Natural Environment Research Council (United Kingdom) through the Oxford Environmental Research Doctoral Training Program and by a Lamb and Flag scholarship from St. John's College, Oxford. The empirical work presented here was supported by National Science Foundation awards EF-062362, DEB-1258231, DEB-1556884, and DEB-0846175.

#### **Statement of Authorship**

T.P. co-conceived the study, developed and performed the analyses, wrote all drafts of the manuscript, and incorporated coauthor suggestions. T.C. co-conceived and supervised the study and interpreted results with T.P. R.D.B., D.N.R., and J.T.-D. established, raised funds for, and ran the mark-recapture experiment. P.B. performed the genotyping and pedigree construction. C.A.H., J.T.-D., E.W.R., and C.K.G. performed the common-garden experiments. All authors contributed to reviewing and editing the final version of the manuscript.

#### **Data and Code Availability**

Data and code underlying the analyses and figures in this article have been deposited in the Dryad Digital Repository (<https://doi.org/10.5061/dryad.jm63xsj7k>; Potter et al. 2020).



## Literature Cited

- Alexander, H. J., J. S. Taylor, S. S. T. Wu, and F. Breden. 2006. Parallel evolution and vicariance in the guppy (*Poecilia reticulata*) over multiple spatial and temporal scales. *Evolution* 60:2352–2369.
- Bassar, R. D., A. López-Sepulcre, D. N. Reznick, and J. Travis. 2013. Experimental evidence for density-dependent regulation and selection on Trinidadian guppy life histories. *American Naturalist* 181:25–38.
- Blyth, C. R. 1972. On Simpson's paradox and the sure-thing principle. *Journal of the American Statistical Association* 67:364–366.
- Bonnet, T., M. B. Morrissey, and L. E. B. Kruuk. 2019. Estimation of genetic variance in fitness, and inference of adaptation, when fitness follows a log-normal distribution. *Journal of Heredity* 110:383–395.
- Bonnet, T., and E. Postma. 2018. Fluctuating selection and its (elusive) evolutionary consequences in a wild rodent population. *Journal of Evolutionary Biology* 31:572–586.
- Bonnet, T., P. Wandeler, G. Camenisch, and E. Postma. 2017. Bigger is fitter? quantitative genetic decomposition of selection reveals an adaptive evolutionary decline of body mass in a wild rodent population. *PLoS Biology* 15:e1002592.
- Charmantier, A., D. Garant, and L. Kruuk. 2014. *Quantitative genetics in the wild*. Oxford University Press, Oxford.
- Cooke, F., P. D. Taylor, C. M. Francis, and R. F. Rockwell. 1990. Directional selection and clutch size in birds. *American Naturalist* 136:261–267.
- Coulson, T., T. G. Benton, P. Lundberg, S. R. X. Dall, and B. E. Kendall. 2006. Putting evolutionary biology back in the ecological theatre: a demographic framework mapping genes to communities. *Evolutionary Ecology Research* 8:1155–1171.
- Coulson, T., T. Potter, and A. Felmy. 2019. Fitness functions, genetic and non-genetic inheritance, and why ecological dynamics and evolution are inevitably linked. *bioRxiv*, <https://doi.org/10.1101/762658>.
- Engen, S., R. Lande, B. E. Sæther, and F. S. Dobson. 2009. Reproductive value and the stochastic demography of age-structured populations. *American Naturalist* 174:795–804.
- Firth, J. A., J. D. Hadfield, A. W. Santure, J. Slate, and B. C. Sheldon. 2015. The influence of nonrandom extra-pair paternity on heritability estimates derived from wild pedigrees. *Evolution* 69:1336–1344.
- Geerts, A. N., J. Vanoverbeke, B. Vanschoenwinkel, W. Van Doorslaer, H. Feuchtmayr, D. Atkinson, B. Moss, T. A. Davidson, C. D. Sayer, and L. De Meester. 2015. Rapid evolution of thermal tolerance in the water flea *Daphnia*. *Nature Climate Change* 5:665–668.
- Gienapp, P., C. Teplitsky, J. S. Alho, J. A. Mills, and J. Merilä. 2008. Climate change and evolution: disentangling environmental and genetic responses. *Molecular Ecology* 17:167–178.
- Grafen, A. 1988. On the uses of data on lifetime reproductive success. Pages 454–471 in T. H. Clutton-Brock, ed. *Reproductive success*. University of Chicago Press, Chicago.
- Grant, P. R., and B. R. Grant. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. *Science* 296:707–711.
- Hadfield, J. D. 2008. Estimating evolutionary parameters when viability selection is operating. *Proceedings of the Royal Society B* 275:723–734.
- . 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software* 33:1–22.
- Hadfield, J. D., A. J. Wilson, D. Garant, B. C. Sheldon, and L. E. Kruuk. 2010. The misuse of BLUP in ecology and evolution. *American Naturalist* 175:116–125.
- Hairston, N. G., and T. A. Dillon. 1990. Fluctuating selection and response in a population of freshwater copepods. *Evolution* 44:1796–1805.
- Handelsman, C. A., E. D. Broder, C. M. Dalton, E. W. Ruell, C. A. Myrick, D. N. Reznick, and C. K. Ghalambor. 2013. Predator-induced phenotypic plasticity in metabolism and rate of growth: rapid adaptation to a novel environment. *Integrative and Comparative Biology* 53:975–988.
- Henderson, C. R. 1950. Estimation of genetic parameters. *Annals of Mathematical Statistics* 21:309–310.
- Hereford, J., T. F. Hansen, and D. Houle. 2004. Comparing strengths of directional selection: how strong is strong? *Evolution* 58:2133–2143.
- Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. *Ecology Letters* 7:1225–1241.
- Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. *American Naturalist* 157:245–261.
- Kruuk, L. E. B. 2004. Estimating genetic parameters in natural populations using the “animal model.” *Philosophical Transactions of the Royal Society B* 359:873–890.
- Kruuk, L. E. B., and J. D. Hadfield. 2007. How to separate genetic and environmental causes of similarity between relatives. *Journal of Evolutionary Biology* 20:1890–1903.
- Kruuk, L. E. B., J. Merilä, and B. C. Sheldon. 2001. Phenotypic selection on a heritable size trait revisited. *American Naturalist* 158:557–571.
- Kruuk, L. E. B., J. Slate, and A. J. Wilson. 2008. New answers for old questions: the evolutionary quantitative genetics of wild animal populations. *Annual Review of Ecology, Evolution, and Systematics* 39:525–548.
- Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–1226.
- Lush, J. L. 1937. *Animal breeding plans*. Iowa State College Press, Ames.
- Lynch, M., and B. Walsh. 1998. *Genetics and analysis of quantitative traits*. Vol. 1. Sinauer, Sunderland, MA.
- Merilä, J., B. C. Sheldon, and L. E. B. Kruuk. 2001. Explaining stasis: microevolutionary studies in natural populations. *Genetica* 112:199–222.
- Morrissey, M. B., L. E. B. Kruuk, and A. J. Wilson. 2010. The danger of applying the breeder's equation in observational studies of natural populations. *Journal of Evolutionary Biology* 23:2277–2288.
- Morrissey, M. B., D. J. Parker, P. Korsten, J. M. Pemberton, L. E. Kruuk, and A. J. Wilson. 2012. The prediction of adaptive evolution: empirical application of the secondary theorem of selection and comparison to the breeder's equation. *Evolution* 66:2399–2410.
- Mrode, R. A. 2014. *Linear models for the prediction of animal breeding values*. CAB, Wallingford.
- Pearl, J. 2014. Understanding Simpson's paradox. *American Statistician* 68:8–13.

- Pemberton, J. M. 2010. Evolution of quantitative traits in the wild: mind the ecology. *Philosophical Transactions of the Royal Society B* 365:2431–2438.
- Pigeon, G., M. Festa-Bianchet, D. W. Coltman, and F. Pelletier. 2016. Intense selective hunting leads to artificial evolution in horn size. *Evolutionary Applications* 9:521–530.
- Pollak, E. J., J. Van der Werf, and R. L. Quaas. 1984. Selection bias and multiple trait evaluation. *Journal of Dairy Science* 67:1590–1595.
- Postma, E. 2006. Implications of the difference between true and predicted breeding values for the study of natural selection and micro-evolution. *Journal of Evolutionary Biology* 19:309–320.
- Potter, T., R. D. Bassar, P. Bentzen, E. W. Ruell, J. Torres-Dowdall, C. A. Handelsman, C. K. Ghalambor, J. Travis, D. N. Reznick, and T. Coulson. 2020. Data from: Environmental change, if unaccounted, prevents detection of cryptic evolution in a wild population. *American Naturalist*, Dryad Digital Repository, <https://doi.org/10.5061/dryad.jm63xsj7k>.
- Potter, T., L. King, J. Travis, and R. Bassar. 2019. Competitive asymmetry and local adaptation in Trinidadian guppies. *Journal of Animal Ecology* 88:330–342.
- Price, G. R. 1970. Selection and covariance. *Nature* 227:520–521.
- Price, T. D., P. R. Grant, H. L. Gibbs, and P. T. Boag. 1984. Recurrent patterns of natural selection in a population of Darwin's finches. *Nature* 309:787–789.
- Pujol, B., S. Blanchet, A. Charmantier, E. Danchin, B. Facon, P. Marrot, F. Roux, I. Scotti, C. Teplitsky, C. E. Thomson, and I. Winney. 2018. The missing response to selection in the wild. *Trends in Ecology and Evolution* 33:337–346.
- Rausher, M. D. 1992. The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. *Evolution* 46:616–626.
- R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>.
- Reznick, D. 1982a. Genetic determination of offspring size in the guppy (*Poecilia reticulata*). *American Naturalist* 120:181–188.
- . 1982b. The impact of predation on life-history evolution in Trinidadian guppies—genetic basis of observed life-history patterns. *Evolution* 36:1236–1250.
- . 1990. Plasticity in age and size at maturity in male guppies (*Poecilia reticulata*): an experimental evaluation of alternative models of development. *Journal of Evolutionary Biology* 3:185–203.
- Reznick, D. N., R. D. Bassar, C. A. Handelsman, C. K. Ghalambor, J. Arendt, T. Coulson, T. Potter, et al. 2019. Eco-evolutionary feedbacks predict the time course of rapid life-history evolution. *American Naturalist* 194:671–692.
- Reznick, D. N., and H. Bryga. 1987. Life-history evolution in guppies (*Poecilia reticulata*). 1. Phenotypic and genetic changes in an introduction experiment. *Evolution* 41:1370–1385.
- Reznick, D. N., M. J. Butler IV, F. H. Rodd, and P. Ross. 1996. Life-history evolution in guppies (*Poecilia reticulata*). 6. Differential mortality as a mechanism for natural selection. *Evolution* 50:1651–1660.
- Reznick, D., and J. A. Endler. 1982. The impact of predation on life-history evolution in Trinidadian guppies (*Poecilia reticulata*). *Evolution* 36:160–177.
- Robertson, A. 1966. A mathematical model of the culling process in dairy cattle. *Animal Science* 8:95–108.
- Ruell, E. W., C. A. Handelsman, C. L. Hawkins, H. R. Sofaer, C. K. Ghalambor, and L. Angeloni. 2013. Fear, food and sexual ornamentation: plasticity of colour development in Trinidadian guppies. *Proceedings of the Royal Society B* 280:20122019.
- Simpson, E. H. 1951. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society B* 13:238–241.
- Smouse, P. E., T. R. Meagher, and C. J. Kobak. 1999. Parentage analysis in *Chamaelirium luteum* (L.) Gray (Liliaceae): why do some males have higher reproductive contributions? *Journal of Evolutionary Biology* 12:1069–1077.
- Stinchcombe, J. R., M. T. Rutter, D. S. Burdick, P. Tiffin, M. D. Rausher, and R. Mauricio. 2002. Testing for environmentally induced bias in phenotypic estimates of natural selection: theory and practice. *American Naturalist* 160:511–523.
- Stopher, K. V., C. A. Walling, A. Morris, F. E. Guinness, T. H. Clutton-Brock, J. M. Pemberton, and D. H. Nussey. 2012. Shared spatial effects on quantitative genetic parameters: accounting for spatial autocorrelation and home range overlap reduces estimates of heritability in wild red deer. *International Journal of Organic Evolution* 66:2411–2426.
- Torres-Dowdall, J., C. A. Handelsman, D. N. Reznick, and C. K. Ghalambor. 2012. Local adaptation and the evolution of phenotypic plasticity in Trinidadian guppies (*Poecilia reticulata*). *Evolution* 66:3432–3443.
- Travis, J., D. N. Reznick, R. D. Bassar, A. Lopez-Sepulcre, R. Ferriere, and T. Coulson. 2014. Do eco-evo feedbacks help us understand nature? answers from studies of the Trinidadian guppy. Pages 1–40 in J. Moya-Larano, J. Rowntree, and G. Woodward, eds. *Eco-evolutionary dynamics*. Vol. 50. *Advances in Ecological Research*. Elsevier, Amsterdam.
- Van Asch, M., L. Salis, L. J. Holleman, B. Van Lith, and M. E. Visser. 2013. Evolutionary response of the egg hatching date of a herbivorous insect under climate change. *Nature Climate Change* 3:244–248.
- Walsh, B., and M. Lynch. 2018. *Evolution and selection of quantitative traits*. Oxford University Press, Oxford.
- Wilson, A. J., D. Réale, M. N. Clements, M. M. Morrissey, E. Postma, C. A. Walling, L. E. B. Kruuk, and D. H. Nussey. 2010. An ecologist's guide to the animal model. *Journal of Animal Ecology* 79:13–26.
- Wood, S. N. 2017. *Generalized additive models: an introduction with R*. Chapman & Hall/CRC, Boca Raton, FL.
- Yule, G. U. 1903. Notes on the theory of association of attributes in statistics. *Biometrika* 2:121–134.

### References Cited Only in the Online Enhancements

- Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis* 1:515–534.
- Jones, O. R., and J. Wang. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources* 10:551–555.
- Reznick, D. N., F. H. Shaw, F. H. Rodd, and R. G. Shaw. 1997. Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). *Science* 275:1934–1937.

Associate Editor: Jill T. Anderson  
Editor: Jennifer A. Lau