

Published in final edited form as:

Nat Ecol Evol. 2018 July ; 2(7): 1093–1103. doi:10.1038/s41559-018-0577-4.

Environmental coupling of heritability and selection is rare and of minor evolutionary significance in wild populations

Jip J.C. Ramakers*, Antica Culina, Marcel E. Visser, and Phillip Gienapp

Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), P.O. Box 50, 6700 AB Wageningen, the Netherlands

Abstract

Predicting the rate of adaptation to environmental change in wild populations is important for understanding evolutionary change. However, predictions may be unreliable if the two key variables affecting the rate of evolutionary change, heritability and selection, are both affected by the same environmental variable. To determine how general such an environmentally induced coupling of heritability and selection is, and how this may influence the rate of adaptation, we made use of freely accessible, open data on pedigreed wild populations to answer this question at the broadest possible scale. Using 16 populations from 10 vertebrate species, which provided data on 50 traits (body mass, morphology, physiology, behaviour and life history), we found evidence for an environmentally induced relationship between heritability and selection in only 6 cases, with weak evidence that this resulted in an increase or decrease in expected selection response. We conclude that such a coupling of heritability and selection is unlikely to strongly affect evolutionary change even though both heritability and selection are commonly postulated to be environment dependent.

Introduction

In the face of global environmental change, it is imperative to understand whether and how fast populations can adapt to novel conditions to be ‘rescued’ by evolution¹. Despite evidence of genetic variance and selection in many wild populations, genetic response to selection (adaptive micro-evolution or rate of adaptation) in natural populations is rarely

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Materials & correspondence: Jip Ramakers, jip.ramakers@gmail.com.

Data availability

Raw data used in the analyses can be found in the references listed in Table 1; DOIs for each dataset can be found in Supplementary Table 1. Data used for the weighted regression analysis, estimating predicted response to selection, and meta-analysis have been uploaded as Supplementary Tables 2 and 3.

Code availability

R code examples for each analysis are available as a supplementary text file.

Author contributions

All authors contributed to the design of the study; A.C. and J.J.C.R. retrieved the datasets; J.J.C.R. and P.G. designed the analysis protocol, with aid from A.C. in the meta-analysis; J.J.C.R. screened the retrieved datasets, contacted the original authors, conducted the analyses and drafted the paper; all authors commented on the manuscript.

Competing financial interests

The authors declare to have no competing financial interests.

observed^{2, 3}. An apparent lack of a response to selection may have a variety of biological and/or methodological causes^{4, 5}. One potential reason is an environmentally induced coupling between selection and additive genetic variation, which can mask the true evolutionary potential of a population if not recognised. By its definition, selection is mediated by the environment^{6, 7} and has been shown to vary from season to season and between geographical regions, depending on resource availability (including mating opportunity) and predation pressure, for example^{8, 9, 10, 11, 12} (but see ref. 13). Genetic variation is, however, also known to vary with the environment (known as genotype-by-environment interaction), being sometimes increased and sometimes reduced under benign conditions (e.g. when mean fitness in the population is high)^{14, 15, 16}, although the ecological drivers of changes in the genetic variance–covariance matrix remain largely unknown¹⁷.

Although the environmental dependency of both selection and genetic variation has been thoroughly documented, our knowledge on how they may interact to result in evolutionary change in natural populations is very limited. In their recent review, Wood and Brodie¹⁶ identified 23 studies that measured environmental effects on selection and 28 studies that measured environmental effects on additive genetic variation. Overall, reviewing a great variety of taxa, environments and traits, they found that environmental effects on selection and genetic variance were broad and inconsistent. Importantly, most studies on environment-dependent genetic variation were done in laboratory settings (and those on selection mostly in wild populations) and extrapolating laboratory findings to natural conditions is not necessarily straightforward. To date, only two studies of natural populations have measured how both genetic variation and selection within the same trait covaried across environments. A study on Soay sheep (*Ovis aries*) demonstrated increased selection for a higher birth weight in harsh environments, whereas total genetic variance was highest in benign environments¹⁸. The opposite was found in the great tit (*Parus major*), where warmer springs, which are associated with increased mismatch between offspring energetic demands and food availability, were associated with stronger selection for early egg-laying as well as high additive genetic variance for that trait¹⁹. Thus, in the former example, selection and genetic variance covaried with the environment in opposite directions, whereas in the latter example they did so in the same direction. The negative covariance between selection and genetic variation in Soay sheep led to a 21% decrease in expected response to selection as opposed to a situation where genetic variance was assumed to not vary with the environment¹⁸. In the great tit, the positive association between additive genetic variance and selection gradients resulted in a 20% increase in predicted response to selection as compared to a situation where heterogeneity in both selection and genetic variance was ignored¹⁹. A more recent study investigated the environmental dependency of genetic variance and selection in several morphological traits in the Soay sheep population, but did not explicitly address the relationship between them, presumably since environment-dependent genetic variance was found to be absent²⁰. The environmental coupling of selection and genetic variance (or heritability) may therefore provide an important explanation for the discrepancy between observed and expected responses to selection in some natural populations, but the prevalence of this mechanism—and how it may alter the expected response to selection—in wild populations remains largely unknown.

We investigated the prevalence and strength of an environmentally induced correlation between heritability and selection—and its expected evolutionary consequence—in a variety of wild populations. We searched for multiannual, pedigreed datasets on wild populations freely accessible from online data repositories and used these data to quantify environment-dependent additive genetic variation (using random regression animal models) and standardised selection gradients for a suite of life-history, morphological, behavioural, physiological and body mass traits. We then regressed heritability against selection for 50 traits from 16 populations and compared expected selection responses with and without considering environmental heterogeneity in heritability. We had no specific expectation as to the prevalence of a correlation between heritability and selection but, if anything, expected it to be more common in life-history and morphological traits, since selection in these traits tends to be strong and variable²¹. Our approach using open data²² speaks to recent recommendations to use available data to address novel, outstanding questions in ecology and evolution that transcend a single study system^{23, 24}.

Results

Data acquisition and author response

We performed a search in online data repositories (see Methods) for multiannual (6 years) datasets containing pedigrees of wild populations accompanied by phenotypic measures on individually marked animals. From 106 acquired pedigreed datasets (Supplementary Table 1), we used 14 that were suitable for our analysis (see Methods). We added one unpublished dataset from our own database (pied flycatcher, *Ficedula hypoleuca*). These 15 datasets comprised 16 different populations, spanning ten species, eight of which were avian species, one a lizard, and one a mammal (Table 1).

Authors were generally supportive of the use of their data. We contacted 14 authors (of 18 datasets) about our use of their data and found that 4 datasets were not usable. This was mainly related to a bias in our approximation of the environment, i.e. the population-mean trait value (see Methods) and selection in a given year when a non-random portion of the population was not represented in the dataset. Only in two cases authors were initially reluctant to cooperate, but all authors eventually informed us about the appropriateness of our analyses of their data (see Culina et al.²⁵ for a full account on author correspondence associated with this article).

Estimating environment, heritability and selection

From the included datasets, we extracted a total of 50 morphological, behavioural, physiological, life-history and body mass traits. We used these traits first to estimate a standardised measure of the environment, the standardised annual population-mean trait value^{26, 27, 28, 29}. We estimated the heritability (h^2 , the relative additive genetic variation) of the traits and found that the majority showed significant heritable variation within the population (Table 1). We then fitted random regression animal models (RRAMs) with an interaction between the additive genetic effect and the standardised measure of the environment. We extracted environment-dependent heritability estimates resulting from these RRAMs (as heritability determines the short-term evolutionary change) and regressed

them against annual standardised selection gradients^{30, 31, 32} (β' ; Supplementary Figure 2), while accounting for uncertainty in both predictor and response. In 6 out of 50 cases (all in bird species), this led to a statistically significant relationship between selection and heritability (Figure 1; Supplementary Table 2). None of the 14 life-history traits exhibited such a relationship, despite considerable variation in both selection and heritability. We found a positive, significant relationship in nestling body mass in *Passerculus sandwichensis*, based on viability selection on nestlings (slope [95% bootstrapped confidence interval] = 0.102 [0.045, 0.191], $r^2 = 0.369$ [0.089, 0.596]). As the only morphological example, nestling tarsus length in *P. sandwichensis* showed a significantly negative correlation based on viability selection (slope = -0.057 [-0.118, -0.023], $r^2 = 0.148$ [0.038, 0.340]). Finally, four avian physiological and behavioural traits exhibited a significant association between heritability and selection, all based on viability selection: *P. major* plumage reflectance at 349 nm (slope = 0.018 [0.009, 0.037], $r^2 = 0.284$ [0.052, 0.507]), 549 nm (slope = -0.190 [-0.440, -0.040], $r^2 = 0.467$ [0.054, 0.949]) and spectral sensitivity (double cone; slope = -0.055 [-0.173, -0.010], $r^2 = 0.248$ [0.027, 0.751]), and *Cyanistes caeruleus* adult handling aggression (slope = 0.001 [0.0004, 0.003], $r^2 = 0.009$ [0.002, 0.021]).

A formal meta-analysis on the correlation coefficient r from each heritability–selection regression, correcting for independence of traits within studies (weighted linear mixed-effects model with random effect ‘study’), reaffirmed that the overall correlation was weak and not dependent on the class of trait (Figure 2). We found similar results when we disregarded non-avian traits.

Comparing expected responses to selection

Environmental coupling of (additive) genetic variance and selection can affect the predicted response to selection^{18, 19}. We therefore predicted for the six datasets identified above the standardised selection response for each environment j (R'_j), assuming either constant or environment-dependent heritability ($R'_j = h^2 \beta'$; or $h_j^2 \beta'_j$). When we calculated the mean difference in response across environments between the two approaches (accounting for uncertainty in estimates), we found that environmental variation in heritability significantly affected the mean expected response in all six cases, but this effect was not in a consistent direction (i.e. either reduced in case of a negative association or increased in case of a positive association; Table 2). Finally, we modelled the directional difference in expected response as a function of the correlation coefficient between h^2 and β' for all datasets (cf. ref. 16), and found that the difference in expected response was not affected by this correlation coefficient (slope = 0.002 [-0.001, 0.004]; Figure 3).

Discussion

Little evidence for environmental coupling of heritability and selection

We investigated the prevalence of an environmentally induced relationship between heritability and selection across traits and study systems by using open data available in data repositories. Our study extends the limited evidence for this phenomenon^{18, 19} to 50 traits

from 10 species in 16 populations. Relying on robust statistical methods to (i) quantify the relationship between heritability and selection, (ii) synthesise results of different studies using meta-analysis, and (iii) infer expected evolutionary response, we conclude that, despite being a current topic in ecology and evolution^{14, 15, 16, 33, 34}, its evolutionary importance in natural populations is small—at least for the range of species for which we have data.

So far, only two studies have investigated this relationship within the same trait and population^{18, 19}. Reanalysis of egg-laying date in the Hoge Veluwe great tit population¹⁹ yielded different results, potentially linked to the different approximation of the environment (see below), although the environment in that population, i.e. mean spring temperature, explains much of the variation in the trait ($r^2 = 0.66$)³⁵. The correlations between selection and heritability or additive genetic variance found by Husby et al.¹⁹, however, were marginally or non-significant, respectively, and were not subjected to rigorous correction for uncertainty like our bootstrapping methods. Thus, even in a population with (i) a strong link between the environment (temperature) and a life-history trait (laying date) and (ii) demonstrated increases in selection and additive genetic variance under increased temperatures, evidence for an environmental link between heritability and selection was modest at best. Heritability of life-history traits is generally found to be low^{26, 36, 37}, potentially due to high environmental variance³⁸ or genetic canalisation³⁹, but life-history traits are inherently likely to exhibit gene-by-environment interactions whenever selection pressures vary with the environment, because of their close association with fitness³⁸. It is, then, remarkable that heritability was not related to selection in any of the life-history traits investigated here (Figure 1), even though substantial variation existed in the strength of selection (Supplementary Figure 2).

Finding a significant relationship between heritability and selection requires sufficient statistical power. Although the number of years and individuals varied considerably between study systems (Table 1), significant relationships were not exclusively found in the largest datasets (Figure 1). A visual inspection of the components that make up this relationship, as well as the relationship between selection and the environment, suggested that statistical significance was neither influenced by the variance in the predictor and response variables nor by the number of years or the total number of observations available (Supplementary Figure 3). Given the larger statistical power associated with larger datasets, the lack of the sought correlation in our largest datasets suggests that the effect size is likely too small to be biologically meaningful.

Using an analytical model informed by data from a literature review, Wood and Brodie¹⁶ predicted that the strength of the relationship between selection and genetic variance would impact the mean and, to a greater degree, variance in responses to selection across hypothetical populations. Yet even in the few cases in which we demonstrated a reasonably strong relationship between heritability and selection (Figure 1; cf. ref. 18), this was not sufficient to fuel a change in the rate of expected response to selection (Table 2). This is partly because both components of the relationship came with prediction error that needed to be accommodated in the estimation of the response. From the studies investigated here, we therefore conclude that even when we find environmental coupling between heritability and selection, its net effect on the predicted evolutionary change is small and is hence an

unlikely explanation for potential discrepancies between observed and expected responses to selection in natural populations⁴.

Methodological considerations

An important aspect in analysis of genotype-by-environment interactions, i.e. testing whether genetic variance and heritability differ among environments, is the choice of the environmental variable. However, in most of the analysed datasets no such environmental variable was included. Instead of obtaining such data from other sources and testing whether the chosen variable was predictive for the trait in question, we used environment-specific, population-mean trait values as the environmental variable (covariate) in our analyses, an accepted practice in animal and plant breeding^{26, 28}. This approach has three major advantages. First, the daunting task of searching environmental data relevant to each trait becomes unnecessary. Second, it enables the inclusion of traits for which it is difficult to conceive and collect environmental data (compare, for example, breeding time in great tits, which is strongly temperature dependent⁴⁰, with a physiological trait like handling aggression in blue tits *C. caeruleus*, for which no clear environmental component has been identified, despite substantial year-to-year and residual variation⁴¹). Third, because the population-mean phenotype encompasses all unmeasured or unobserved components of the environment, it will generally be an accurate representation of the environment for the trait of interest²⁶, circumventing the problem of misidentifying the relevant environmental component and, consequently, erroneously inferring the presence or absence of variation in reaction norm slopes. For example, in a population of collared flycatchers (*Ficedula albicollis*), Brommer et al.⁴² found significant between-individual variation in breeding-time reaction norms in response to average temperatures in spring, but not to rainfall or North Atlantic Oscillation, even though these variables correlated well with breeding time. Similarly, Husby et al.⁴³ could show between-individual variation in reaction norms for breeding time in great tits while Charmantier et al.⁴⁴ did not find this in the same population when using a different environmental variable. Indeed, simulations have shown that random regression models with ‘mean trait’ as the environment yielded similar variation in reaction norm slopes to models with a ‘real’ environmental driver of the trait⁴⁵. In contrast, using other environmental measures that did not drive the trait but correlated with the ‘real’ environment to a decreasing degree ($r = 0.9$ to 0.1) yielded increasingly downwardly biased estimates of both the slope and the variance in the reaction norm. This is an important finding because it shows that environment-specific mean phenotypes can serve as a ‘yardstick’ when testing for gene-by-environment interactions⁴⁵.

Ideally, heritability should be estimated at the same level as where selection operates, because the correlation at this level is what ultimately matters. Since selection is generally estimated at an annual level (where each year captures all components of the environment), heritability should be estimated at this level too. This would, however, require an enormous number of individuals in each year to estimate the annual genetic variances reliably—and hence generally not be feasible. Using (continuous) environmental covariates instead to estimate genetic variance along an environmental gradient^{46, 47} is the next best option, and the best way to do this is to choose a metric that captures most features of the environment in a given year (which annual mean phenotypes do). This alleviates the need to establish a

link between an environmental covariate and selection, which will not necessarily be informative when investigating the correlation between heritability and selection—in particular when statistical power is limited.

A concern when estimating selection in natural populations is to identify the real target of selection^{30, 32, 48}. The use of the Breeders' Equation to predict evolutionary change in natural populations has therefore been advised against, and the Robertson-Price identity has been suggested as an appropriate alternative^{48, 49}. However, estimating the genetic covariance between a trait and fitness at an annual basis to estimate variation in selection is rarely, if ever, possible, due to the large datasets required to reliably estimate genetic covariances. Furthermore, Reed et al.⁵⁰ showed that in a wild population of great tits, environmental bias in phenotypic selection estimates for egg-laying date and clutch size is small at best. A similar conclusion was reached by Morrissey and Ferguson⁵¹, who showed for brook charr (*Salvelinus fontinalis*) that estimates of phenotypic selection on body size are highly congruent with estimates of genetic selection.

Benefits and limitations of open data

One important development in ecology and evolution in recent years has been the requirement to make the data used to produce the results of studies (usually published studies) publicly available^{23, 52}, leading to a surge in data output onto online data repositories. The potential advantages of open data archiving in revolutionising the natural sciences are now increasingly recognised^{22, 24}. Yet Evans⁵³ showed that data from long-term population studies archived in Dryad Digital Repository are never used by third parties. Our multi-study approach makes extensive use of such long-term data to address an outstanding question in evolutionary ecology. Indeed, the use of open data comes with important logistical and ethical issues^{52, 54, 55} that need to be addressed before biological conclusions can be safely drawn. Our study, however, shows that it can be done successfully (see also ref. 25).

From the 106 initially considered datasets in our example, we could eventually use only 14 (plus the previously unpublished pied flycatcher dataset), due to various reasons such as small and/or biased sample sizes, a lack of appropriate fitness data, unusable pedigrees (e.g. relatedness matrices, which we were unable to use after data manipulation because they required a specific ordering of the individuals in the phenotype file), and a low number of years. Moreover, the data were heavily biased towards birds and mammals (50 and 31 datasets, respectively). We therefore need to make the cautionary note that we cannot necessarily extrapolate the evolutionary importance of an environmental correlation between selection and genetic variance across a wider range of taxa. The general taxon bias in quantitative genetic studies of wild populations toward birds and mammals can be explained by the fact that linking individual offspring to their parents, necessary to construct a pedigree, is comparably straightforward⁵⁶. Relatedness matrices based on genomic markers may make pairwise relatedness estimates a less stringent requirement in evolutionary studies in the future and in that way greatly augment the taxonomic scale at which important evolutionary questions can be addressed⁵⁷. Time will resolve issues like samples sizes and number of years, but whether or not a dataset is suitable will ultimately depend on the type

of analysis and the type of data required. In the era of Open Science that encourages publication of datasets while increasing their quality, it is but a matter of time before taxon biases in multi-annual meta-studies similar to ours may dissipate. Such long-term datasets of individually marked animals are invaluable tools in ecology and evolution and will inevitably serve to elucidate the ecological and evolutionary consequences of environmental change^{56, 58}.

Methods

Data acquisition

In May and July 2016 we conducted a search for datasets that contained pedigree information on a wild species through twelve different aggregators of research data repositories (Europe PMC, DataCite, BASE, OpenAIRE, Science Research, DataOne Mercury search, Web of Science Data Citation Index, Scielo, Research Data Australia, DLI Service, Dryad Digital Repository, DataMED). These aggregators collect information on datasets (e.g. title, keywords, abstract and description) that have been deposited in different data repositories, and allow for search through multiple data sources in one search interface. Datasets were tracked using fixed search terms (see Supplementary Methods 1); search results were screened based on title, abstract, dataset description, and/or keywords, if available. Remaining datasets were further checked for relevance by opening the data files and/or reading the related publication if necessary, leaving only datasets containing pedigree information for a wild or captive animal population. Recording of datasets was done according to PRISMA guidelines^{25, 59}.

Next, we screened and filtered this data subset (103 datasets) to keep those where: (i) the pedigree file could potentially be used (i.e. when the file was not embargoed, corrupted or otherwise unsuitable for our particular analysis, e.g. relatedness matrices lacking the specific links between parents and offspring); (ii) the pedigree contained a sufficient number of individuals (final datasets had, on average, >40 observations/individuals per year); (iii) individuals in the pedigree also had information on a phenotype on which selection could act; (iv) there was natural environmental variation in the phenotype (this excluded all laboratory populations); (v) the associated phenotype file contained at least six years of data; and (vi) there were no additional issues (e.g. non-matching IDs of animals in pedigree and phenotype file). In addition to these 103 datasets, we did an additional search in Web of Science (on 9 September 2017; see Supplementary Methods 1) and from the resulting 396 studies, we discovered three additional suitable datasets overlooked by the initial search, using the inclusion criteria above. Lastly, we added our own, previously unpublished data from the long-term study of pied flycatchers (*Ficedula hypoleuca*, ref. 60; see ref. 61 for more information on that population), totalling 107 retrieved datasets (Supplementary Table 1).

The total number of datasets included in the analysis amounted to 15, covering 10 species from 16 populations and a variety of life-history, morphological, physiological, behavioural and body mass traits (Table 1). This excludes datasets that initially appeared suitable to us but whose suitability for our analysis was refuted by the original authors (see 'Enquiring with original authors'; Supplementary Table 1).

Quantifying the environment

None but two of the final datasets provided information about the environment. Therefore, we used a standardised protocol to quantify the environment. For each year, we calculated the population-mean trait value (\bar{x}) as a measure of the general environment and mean- and variance-standardized it across seasons/sites:

$$E'_j = \frac{\bar{x}_j - \mu_{\bar{x}}}{\sigma_{\bar{x}}},$$

where j denotes the j^{th} season, and μ and σ the grand mean and standard deviation, respectively. Note that this measure does not identify any specific environmental parameter but captures the environment as a whole in a specific season. The method is commonly used in animal and plant breeding studies in a process called ‘joint-regression analysis’, where genotype-specific interactions are partitioned into a component explained by mean population performance and a residual component (ref. 26, pp. 672–678). It was first proposed by Yates and Cochran²⁹ and later brought into prominence in a barley yield experiment by Finlay and Wilkinson²⁷, and has now become widely accepted in the plant- and animal-breeding literature^{26, 28}. It has the advantage that all of the complex (and potentially unobserved) features of the environment are integrated into a single measure, allowing for the ranking of seasons in terms of overall environmental quality. Note that this method disqualifies traits that do not vary at the annual level (i.e. fixed adult traits were not used in our analyses).

One complication with our measure of the environment is that such a measure is potentially biased when a non-random portion of the population in a given season is removed from the dataset (e.g. because certain individuals are never sampled), or when changes in the demographic structure of the population strongly affect the mean trait value. When this was the case (see ‘Enquiring with original authors’; Supplementary Table 1), the dataset was dropped from further analysis.

Standard trait heritability

For each of the traits in our full data (Table 1), we tested for evidence of additive genetic variance following a standardized protocol. First, we constructed ‘minimum adequate’ mixed-effects models (MAMs) with the trait of interest as response variable (all with Gaussian errors) using restricted maximum likelihood (REML) estimation in ASReml-R^{62, 63}. This method provides a fast and efficient way of estimating variance components and allows for the inclusion of additive genetic effects. Note, however, that we used a Bayesian approach to estimate environment-dependent heritability estimates, as this allows for estimation of posterior confidence regions, which we needed to reliably account for uncertainty in our environment-dependent heritability estimates in subsequent analysis (see ‘Genotype-by-environment analysis’). Fixed effects were the environment (E'), as continuous variable, and additional effects provided in the dataset, based on mixed-effects models in the associated original paper. Significance of these effects, as well as that of interactions between effects, was tested with conditional Wald F tests, removing non-significant ($p > 0.05$) terms in a backward stepwise manner (but always keeping E').

Random effects were those identified in the original papers (always containing a ‘permanent environment’ effect, i.e. individual ID, when there were multiple observations of the same individual), but sometimes we constructed our own additional effects when deemed biologically appropriate (e.g. in nestling traits, ‘nest-box ID’ and ‘year’ were combined to identify common-environment effects within a single brood). Significance of random effects was tested using likelihood-ratio tests ($D = 2[\log(L_{ml}) - \log(L_{m0})]$, where D is asymptotically χ^2 distributed with one degree of freedom). The MAM was extended to an ‘animal model’^{64, 65} (AM) by adding a random additive genetic effect based on the pedigree with maternal and paternal links (see references in Table 1 for how pedigrees were constructed). Thus, the AMs took the form

$$y = \mu + X_1\beta_{E'} + \dots + X_n\beta + Z_1pe + Z_2a + \dots + Z_nu + \varepsilon,$$

where y is a vector of phenotypes, $X_{1\dots n}$ and $Z_{1\dots n}$ are the design matrices relating the fixed (β) and random effects (pe , permanent environment; a , additive genetic; u , other) to y , μ is the mean and ε is the error term. The narrow-sense heritability was calculated as $h^2 = \sigma_a^2 / \sigma_p^2$, where σ_p^2 represents the total phenotypic variance comprising all variance components, conditioned on the fixed effects (Table 1).

Genotype-by-environment analysis

To model the interaction between additive genetic variance and the environment (G×E), we extended the AM to a random regression animal model (RRAM) using the ‘MCMCglmm’ package^{66, 67} (ignoring years with <8 observations). In the RRAMs, we allowed the environment to interact with both the permanent environment (if present) and the additive genetic effect:

$$y = \mu + X_1\beta_{E'} + \dots + X_n\beta + Z_1(pe, E', n_1) + Z_2(a, E', n_1) + \dots + Z_nu + \varepsilon,$$

where n_1 is the first-order polynomial of the regression function. Fixed and random terms were those identified from the (M)AMs; note that because E' explains most of the variation related to seasonal effects, it replaced the random effect of year in most analyses. We constructed two 2×2 unstructured variance–covariance matrices for the intercept and the slope of the permanent environment and the additive genetic effect:

$$\mathbf{P} = \begin{bmatrix} \sigma_{pe_{int}}^2 & \sigma_{pe_{int}, pe_{sl}} \\ \sigma_{pe_{sl}, pe_{int}} & \sigma_{pe_{sl}}^2 \end{bmatrix} \text{ and } \mathbf{G} = \begin{bmatrix} \sigma_{a_{int}}^2 & \sigma_{a_{int}, a_{sl}} \\ \sigma_{a_{sl}, a_{int}} & \sigma_{a_{sl}}^2 \end{bmatrix}.$$

In cases where there was no permanent-environment effect but only a maternal or common-environment effect (in juvenile-only traits), only the \mathbf{G} matrix was fitted. To avoid artificial inflation of slope variance estimates in the \mathbf{P} and \mathbf{G} matrices due to heterogeneity in residual variance across the environmental gradient, we partitioned the residual component ε into

‘environmental blocks’⁶⁸, following categorisation of environments into n equal-interval groups. Thus, we fitted the residual matrix as an $n \times n$ identity matrix,

$$\mathbf{R} = \begin{bmatrix} \sigma_{\varepsilon_1}^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{\varepsilon_n}^2 \end{bmatrix},$$

where n was the number of environments divided by 5, but was always ≥ 3 (e.g. in a dataset with 20 environments $n = 4$, but with 10 environments $n = 3$).

To illustrate that our environmental metric (E') was valid in this context, Gienapp⁴⁵ ran random regression models on simulated data using several different quantifications of the environment related to a ‘true’ environmental driver of the phenotype, as well as the annual trait mean. He found no evidence that variance estimates of reaction norm intercepts and slopes were biased by the annual trait mean (relative to the ‘true’ environmental driver) and showed that this metric outperformed environmental correlates. Although we concur that additive genetic variance may not only be affected by current environmental conditions but also be the outcome of past selection processes, it is evident from many quantitative genetic studies of wild populations that year-to-year variation in phenotypes is mostly attributable to phenotypic plasticity and that the share of genetic change from year to year is generally very small and undetectable^{2, 3}. Consequently, we believe that using environment-specific mean trait values will lead to more reliable results than an environmental variable that correlates too weakly with the real driver of plasticity, thereby underestimating variation in (genetic) reaction norm slopes.

To obtain independent samples in the MCMC sampling process, we used a thinning interval of 20,000 in all models, with a burn-in period of 200,000 samples and a total effective sample size of 250 (i.e. 5,200,000 samples). In exploratory stages of the analysis, we found that a larger effective sample size (1000) did not affect the posterior estimates, but these models take substantially longer to complete. Effective sample size in all models included never fell substantially below 250 and autocorrelation between samples was almost always < 0.1 but never exceeded 0.2 for any variance component; models that did not meet these criteria were discarded (not listed in Table 1). For the residual term, we specified Inverse-Wishart (IW) priors ($\mathbf{V} = \text{diag}(x)$ and $\text{nu} = 1.002$, where x is the dimension of the matrix). For the random terms we explored two alternative priors: the IW prior (specifications as above) and parameter-expanded (PE) priors ($\mathbf{V} = \text{diag}(x)$, $\text{nu} = x$, $\text{alpha} \cdot \mu = 0$, $\text{alpha} \cdot \mathbf{V} = \text{diag}(x) \cdot 500$). Although both priors yielded similar results in most cases, the posterior variances tended to be smaller when real variance was close to zero under the PE compared to the IW prior. This is in agreement with previously voiced concerns that the IW prior may behave poorly when true variance is close to zero^{66, 69, 70}. We therefore only present posterior estimates from the models based on PE priors. We refrained from ‘significance’ testing of the $G \times E$ interaction, because—issues concerning model-selection criteria such as DIC aside^{66, 71, 72}—of potentially limited power in the smaller datasets and our main interest in testing the covariance between h^2 and selection. We instead opted for a pragmatic

approach and used the highest posterior density intervals (HPDIs) to account for uncertainty in all subsequent analyses. The rationale behind this was that if we had excluded all ‘non-significant’ G×E interactions, of which some may have been false negatives, we may have overlooked a potentially strong covariance between h^2 and selection (see below). By accounting for the uncertainties in environment-specific h^2 estimates, the true negatives in G×E will not lead to a spurious covariance between h^2 and selection.

The posterior mean variance for each variance component in environment j was derived from the estimated **G** and **P** matrix as⁷³

$$\sigma_j^2 = \sigma_{int}^2 + 2\sigma_{int,sl}\beta_{E'j} + \sigma_{sl}^2\beta_{E'j}^2.$$

The 95% HPDIs were likewise derived from the upper and lower HPDI matrices. Environment-dependent heritability was defined as the mean of posterior variance estimates, $h_j^2 = \sigma_{a_j}^2 / (\sigma_{a_j}^2 + \sigma_{pe_j}^2 + \dots + \sigma_{\epsilon_j}^2)$, with 95% HPDIs estimated from the lower and upper HPD limits of each variance component. Standard errors of h_j^2 were then calculated as half the 95% HPDI divided by 1.96.

Selection

To quantify selection on the trait in a given environment, we made use of provided reproductive fitness data (number of offspring or recruits) or survival data (Table 1). When such data were not provided, we inferred (annual) reproductive success by linking animals to sires and dams in the pedigree using their birth year (when available). If we could not infer annual recruits from the pedigree, we determined survival from one year to the next by identifying reappearance of individuals in the dataset in subsequent years, assuming the last year of appearance was the last year the individual was alive. As with quantifying the environment (see above), inferring fitness is problematic if a non-random portion of the population appears in the dataset (aside from the non-random disappearance due to selection; see also ref. 48). When this was likely to be problematic (see ‘Enquiring with original authors’; Supplementary Table 1) the dataset in question forewent inclusion in the analysis.

To estimate annual, standardised selection gradients (β'), we constructed general(ised) additive models (GAMs, package ‘mgcv’⁷⁴), where the fitness component was the response variable following either a Gaussian, Poisson or negative binomial distribution for fecundity measures (number of offspring produced or recruits), depending on the distribution of the data, or a binomial distribution for survival (1/0 response). As fixed effects, we initially included an interaction between year and the trait of interest and used it as a null model to identify additional significant fixed effects (using F or χ^2 tests) that influenced the fitness measure (e.g. age or sex and additional quantitative traits). Based on these findings, we ran annual GAMs (without ‘year’) and calculated annual β' using the ‘gam.gradients’ function from the ‘gsg’ package⁷⁵. This procedure estimates β' s as

$$\beta' = \frac{\text{Cov}(w, z)}{\sigma_z},$$

where the numerator is the covariance between the trait and relative fitness, i.e. the partial regression coefficient after taking into account the effect of traits potentially simultaneously under selection, and the denominator is the standard deviation of the trait, following Lande and Arnold^{30, 31}. Standard errors of β' were estimated through parametric bootstrapping (1000 iterations).

Covariance between selection and heritability: a meta-analysis

As we were interested in studying the effect of environmental variation in selection and genetic variance on selection response, we examined the (linear) relationship between heritability and selection. We refrained from making this analysis conditional on the presence of an underlying correlation between β' and E , because in cases where statistical power may be an issue, such a two-step approach would decrease the likelihood of detecting a real relationship between h^2 and β' if datasets were omitted based on this criterion. A similar reasoning applied to testing for an underlying relationship between h^2 and E (see above). For each dataset, we regressed h^2 against β' in linear weighted least-squares (WLS) regressions, weighting data points by $1/[(\text{standard error of } h^2)^2]$. To account for uncertainty in the predictor, β' , we substituted each of its values (j) with a randomly drawn value from a random normal distribution ($n = 1000$, $\mu = \beta'_j$ and $\sigma = \text{standard error of } \beta'_j$) and iterated the entire process 1000 times. We obtained the mean and the 0.025 and 0.975 quantiles (i.e. the 95% bootstrapped confidence interval CI) of the model estimates (intercepts and slopes) resulting from these iterations; estimates were considered statistically significant if the 95% CI did not include 0. Note that in reality, estimates of β' are not entirely independent because some individuals are included in multiple estimates, potentially affecting the estimates from (W)LS regression models. We believe, however, that this issue was sufficiently accounted for by our pragmatic bootstrapping approach.

When estimating the covariance between selection and heritability we took the sign of the estimated selection gradients into account, i.e. we did not correlate heritability with the *absolute* strength of selection (cf. ref. 16). The rationale was that (1) it is biologically relevant whether there is selection for larger or smaller trait values and (2) using absolute or signed selection gradients has different implications for evolutionary change. If a correlation between absolute strength of selection and heritability exists, the overall selection response will not be altered because episodes of strong selection in either direction are always coupled with either high or low heritability. This is, however, not the case when signed selection estimates are used, because in this case strong selection in one direction is coupled with low heritability, whereas strong selection in the other direction is coupled with high heritability.

To examine the overall correlation coefficient across studies and trait types, we performed a meta-analysis using the mean correlation coefficients (r) and their standard errors (SE_r , i.e. half the 95% CI divided by 1.96) resulting from each bootstrapped regression model.

Following Nakagawa and Cuthill76, we transformed coefficients prior to meta-analysis to Fisher's Z_r ,

$$Z_r = 0.5 \times \ln\left(\frac{1+r}{1-r}\right).$$

Variance in Z_r was calculated as77

$$\sigma_{Z_r}^2 = SE_r^2 \times \left(\frac{1}{[1+r] \times [1-r]}\right)^2.$$

We estimated the (weighted) mean correlation coefficient ($n = 50$) in a linear mixed-effects model (REML, package 'lme4'78) with trait type (life history, body mass, morphology, or other) as a fixed effect, study area (i.e. by species; $n = 16$) as a random effect, and $1/\sigma_{Z_r}^2$ as weights. We initially included a random effect of species, which explained 0 variance and was therefore removed from the model (note that the bias toward passerine birds in the acquired datasets precluded phylogenetic analysis). Mean Z_r and 95% CI, predictions unconditioned on the random term, were calculated for each trait type and from a null model excluding the fixed term (i.e. intercept only) through bootstrapping with 1000 iterations. The procedure was repeated on a subset of the data that excluded non-avian traits ($n = 43$ coefficients, 14 studies). To quantify the consistency among studies, we estimated for both sets of analysis (all data or avian-only) the heterogeneity (I^2 , the proportion of variance that cannot be explained by chance) in the random-effects components for the random-only models (see ref. 79 for details). Residual variance in Z_r was estimated at 1.55 and 1.76, respectively, whereas 'study' variance was 0.002 in both cases. Error variance (σ_m^2 in ref. 79) was small (0.016 and 0.014, respectively) and I^2 was estimated at 0.99 in both cases.

Expected response to selection

To quantify the consequence of a covariance between h^2 and β' on the response to selection, we predicted the absolute response to selection under the assumption of constant vs varying heritability following the Breeder's Equation30, 80, i.e. $R'_j = R_j \sigma_z^{-1} = h^2 \beta'_j$ vs. $h_j^2 \beta'_j$. Note that the expected response is in units standard deviation30, 32 (hence σ_z), indicated by the apostrophe. The standard error for R'_j was derived by adding up the relative standard errors of h^2 (or h_j^2) and β'_j . We then calculated the mean absolute (1) and directional (2) difference in response between the two approximations (R'), with the assumption that non-constant heritability does affect the response from any one year to the next (1) and that this difference is directional (2), i.e. positive when the correlation between h^2 and β' is positive and vice versa18, 19. We estimated mean R' across seasons in a linear model without an intercept and with a fixed effect of 'study'. As a response variable, R' in each environment (j) was determined as the difference between two randomly drawn (absolute) values for R'_j from two random normal distributions ($n = 1000$, $\mu = R'_j$ and $\sigma =$ standard error of R'_j). Mean R' was

derived as the mean, study-specific intercept from 1000 iterations, along with the 0.025 and 0.975 quantiles (i.e. 95% CI).

Loosely based on Wood and Brodie III16, we estimated whether the strength of the relationship between heritability and selection affected expected (difference in) selection response. We repeated the procedure above for all the datasets (except those for which $h^2 = 0$) and calculated the expected, mean directional difference (\pm standard error) in expected response to selection assuming varying *vs.* constant heritability (R'). We also extracted the correlation coefficients, r , along with their 95% CIs, from each WLS regression model described in the previous section and calculated standard errors of r as half the 95% CI divided by 1.96. We ran a WLS regression model with R' as a response variable and $1/[(\text{standard error of } R')^2]$ as weights. The correlation coefficient r was the predictor, randomly drawn from a random normal distribution ($n = 1000$, $\mu = r$ and $\sigma = \text{standard error of } r$); the procedure was iterated 1000 times and mean estimates and the 0.025 and 0.975 quantiles (95% CI) were extracted. We also tested this relationship with 'study area' as a random effect in a linear mixed-effects model, but found that this factor explained 0 variance.

Enquiring with original authors

A potential danger of using open data is that the investigator may not be familiar with the study system and therefore make false assumptions about the data^{52, 54, 55}. Hence, for every dataset potentially suitable for analysis, we wrote a letter to the leading author and/or principal investigator of the associated paper, informing them about the general project aim, as well as a description with specifics regarding the use of their dataset (see Supplementary Methods 2). The description contained information about which data files we used, what our study aim was using their datasets, how we went about preparing the data for analysis (e.g. combining multiple files, (re)construction of the pedigree, calculation of the environment based on the population-mean trait value, identification of reproductive performance or survival), how we analysed the data (including which variables we included in the (M)AMs and RRAMs) and a brief overview of tentative findings. We were specifically interested in the authors' verdict on our quantification of the environment and fitness. All analyses presented here are based on datasets that were deemed 'appropriately used' by the original authors. A common concern with discarded datasets was that reproductive success or survival could not be reliably inferred, for example because a non-random portion of recruits disperse away from the study area, or because surviving individuals were not included in the dataset because they had no phenotype. Similarly, non-random dropping of individuals was likely to affect the estimation of the environment (E'), in which case the dataset forewent inclusion in the analysis. We refer the reader to Supplementary Table 1 for a full list of considered datasets and the reason for their exclusion. We report on the author correspondence in more detail in Culina et al.²⁵.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to all the original data owners who found time in their tight schedules to reply to our e-mails, or who otherwise contributed to the discussion (these include people whose data we did not end up using for our analyses). Specifically, we thank—in no particular order—Ben Sheldon, Arild Husby, Erik Postma, Brigitte Schloegl, Katharina Foerster, Juan Fargallo, Jon Brommer, Barbara Class, Anders Møller, Gabriella Ljungstrom, Simon Blanchet, Nat Wheelwright, Marion Nicolaus, Niels Dingemanse, Scott Sakaluk, Alastair Wilson, Jarrod Hadfield, Joey Hubbard, Jane Reid, Peter Arcese, Elise Huchard, Tim Clutton-Brock, Josephine Pemberton, Susan Johnston, Timothée Bonnet, and Boris Delahaie. Shinichi Nakagawa kindly advised us on the meta-analysis. Thomas Reed, Arild Husby and three anonymous reviewers provided useful comments on our manuscript. This work was funded in part by NIOO Strategic funds (SM1521) and an ERC Advanced Grant (339092 - E-Response) to M.E.V.

References

1. Carlson SM, Cunningham CJ, Westley PAH. Evolutionary rescue in a changing world. *Trends in Ecology & Evolution*. 2014; 29:521–530. [PubMed: 25038023]
2. Gienapp P, Teplitsky C, Alho JS, Mills JA, Merilä J. Climate change and evolution: disentangling environmental and genetic responses. *Molecular Ecology*. 2008; 17:167–178. [PubMed: 18173499]
3. Merilä J, Hendry AP. Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evolutionary Applications*. 2014; 7:1–14. [PubMed: 24454544]
4. Merilä J, Sheldon BC, Kruuk LEB. Explaining stasis: microevolutionary studies in natural populations. *Genetica*. 2001; 112–113:199–222.
5. Kruuk LEB, Merilä J, Sheldon BC. When environmental variation short-circuits natural selection. *Trends in Ecology & Evolution*. 2003; 18:207–209.
6. Darwin, C. *On the Origin of Species by Means of Natural Selection*. John Murray; 1859.
7. Wade MJ, Kalisz S. The causes of natural selection. *Evolution*. 1990; 44:1947–1955. [PubMed: 28564434]
8. Weese DJ, Gordon SP, Hendry AP, Kinnison MT. Spatiotemporal variation in linear natural selection on body color in wild guppies (*Poecilia reticulata*). *Evolution*. 2010; 64:1802–1815. [PubMed: 20067520]
9. Gosden TP, Svensson EI. Spatial and temporal dynamics in a sexual selection mosaic. *Evolution*. 2008; 62:845–856. [PubMed: 18194470]
10. Grant PR, Grant BR. Unpredictable evolution in a 30-year study of Darwin's finches. *Science*. 2002; 296:707–711. [PubMed: 11976447]
11. Hairston NG, Dillon TA. Fluctuating selection and response in a population of freshwater copepods. *Evolution*. 1990; 44:1796–1805. [PubMed: 28567814]
12. Siepielski AM, DiBattista D, Carlson SM. It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecology Letters*. 2009; 12:1261–1276. [PubMed: 19740111]
13. Morrissey MB, Hadfield JD. Directional selection in temporally replicated studies is remarkably consistent. *Evolution*. 2012; 66:435–442. [PubMed: 22276539]
14. Hoffman AA, Merilä J. Heritable variation and evolution under favourable and unfavourable conditions. *Trends in Ecology & Evolution*. 1999; 14:96–101. [PubMed: 10322508]
15. Lédon-Rettig CC, Pfennig DW, Chunco AJ, Dworkin I. Cryptic genetic variation in natural populations: a predictive framework. *Integrative & Comparative Biology*. 2014; 54:783–793. [PubMed: 24944116]
16. Wood CW, Brodie ED III. Evolutionary response when selection and genetic variation covary across environments. *Ecology Letters*. 2016; 19:1189–1200. [PubMed: 27531600]
17. Wood CW, Brodie ED. Environmental effects on the structure of the G-matrix. *Evolution*. 2015; 69:2927–2940. [PubMed: 26462609]
18. Wilson AJ, et al. Environmental coupling of selection and heritability limits evolution. *PLoS Biology*. 2006; 4:e216. [PubMed: 16756391]
19. Husby A, Visser ME, Kruuk LEB. Speeding up microevolution: the effects of increasing temperature on selection and genetic variance in a wild bird population. *PLoS Biology*. 2011; 9:e1000585. [PubMed: 21408101]

20. Hayward AD, Pemberton JM, Berenos C, Wilson AJ, Pilkington JG, Kruuk LE. Evidence for Selection-by-Environment but Not Genotype-by-Environment Interactions for Fitness-Related Traits in a Wild Mammal Population. *Genetics*. 2018; 208:349–364. [PubMed: 29127262]
21. Kingsolver JG, et al. The strength of phenotypic selection in natural populations. *The American Naturalist*. 2001; 157:245–261.
22. Culina A, Baglioni M, Crowther TW, Visser ME, Woutersen S, Manghi P. Navigating the unfolding open data landscape in ecology and evolution. *Nature Ecology & Evolution*. 2018; 2:420–426. [PubMed: 29453350]
23. Whitlock MC, McPeck MA, Rausher MD, Rieseberg L, Moore AJ. Data archiving. *American Naturalist*. 2010; 175:145–146.
24. Hampton SE, et al. Big data and the future of ecology. *Frontiers in Ecology and the Environment*. 2013; 11:156–162.
25. Culina A, Crowther TW, Ramakers JJC, Gienapp P, Visser ME. How to do meta-analysis of open datasets. *Nature Ecology & Evolution*. (in press).
26. Lynch, M., Walsh, B. *Genetics and analysis of quantitative traits*. Sinauer Associates, Inc.; 1998.
27. Finlay KW, Wilkinson GN. The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research*. 1963; 14:742–754.
28. James, JW. Genotype by environment interaction in farm animals. *Adaptation and fitness in animal populations: Evolutionary and breeding perspectives on genetic resource management*. van der Werf, J.Graser, H-U.Frankham, R., Gondro, C., editors. Springer; 2009.
29. Yates F, Cochran WG. The analysis of groups of experiments. *Journal of Agricultural Science*. 1938; 28:556–580.
30. Lande R, Arnold SJ. The measurement of selection on correlated characters. *Evolution*. 1983; 37:1210–1226. [PubMed: 28556011]
31. Morrissey MB, Sakrejda K. Unification of regression-based methods for the analysis of natural selection. *Evolution*. 2013; 67:2094–2100. [PubMed: 23815662]
32. Hereford J, Hansen TF, Houle D. Comparing strengths of directional selection: how strong is strong? *Evolution*. 2004; 58:2133–2143. [PubMed: 15562680]
33. Garant D, Kruuk LEB, Wilkin TA, McCleery RH, Sheldon BC. Evolution driven by differential dispersal within a wild bird population. *Nature*. 2005; 433:60–65. [PubMed: 15635409]
34. Garant D, Kruuk LEB, McCleery RH, Sheldon BC. Evolution in a changing environment: a case study with great tit fledging mass. *The American Naturalist*. 2004; 164:E115–E129.
35. Visser ME, Holleman LJM, Gienapp P. Shifts in caterpillar biomass phenology due to climate change and its impact on the breeding biology of an insectivorous bird. *Oecologia*. 2006; 147:164–172. [PubMed: 16328547]
36. Charmantier, A., Garant, D., Kruuk, LEB. *Quantitative genetics in the wild*. Oxford University Press; 2014.
37. Houle D. Comparing evolvability and variability of quantitative traits. *Genetics*. 1992; 130:195–204. [PubMed: 1732160]
38. Price T, Schluter D. On the low heritability of life-history traits. *Evolution*. 1991; 45:853–861. [PubMed: 28564058]
39. Stearns SC, Kawecki TJ. Fitness sensitivity and the canalization of life-history traits. *Evolution*. 1994; 48:1438–1450. [PubMed: 28568425]
40. Visser ME, Holleman LJM, Caro SP. Temperature has a causal effect on avian timing of reproduction. *Proc R Soc Lond B*. 2009; 276:2323–2331.
41. Class B, Brommer JE. A strong genetic correlation underlying a behavioural syndrome disappears during development because of genotype-age interactions. *Proceedings of the Royal Society B-Biological Sciences*. 2015; 282 20142777.
42. Brommer JE, Merilä J, Sheldon BC, Gustafsson L. Natural selection and genetic variation for reproductive reaction norms in a wild bird population. *Evolution*. 2005; 59:1362–1371. [PubMed: 16050111]

43. Husby A, Nussey DH, Visser ME, Wilson AJ, Sheldon BC, Kruuk LEB. Contrasting patterns of phenotypic plasticity in reproductive traits in two great tit (*Parus major*) populations. *Evolution*. 2010; 64:2221–2237. [PubMed: 20298465]
44. Charmantier A, McCleery RH, Cole LR, Perrins CM, Kruuk LEB, Sheldon BC. Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science*. 2008; 320:800–803. [PubMed: 18467590]
45. Gienapp P. The choice of the environmental covariate affects the power to detect individual variation in reaction norm slopes. *BioRxiv*. 2018; doi: 10.1101/311217
46. Schaeffer LR. Application of random regression models in animal breeding. *Livestock Production Science*. 2004; 86:35–45.
47. Nussey DH, Wilson AJ, Brommer JE. The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology*. 2007; 20:831–844. [PubMed: 17465894]
48. Hadfield J. Estimating evolutionary parameters when viability selection is operating. *Proceedings of the Royal Society B-Biological Sciences*. 2008; 275:723–734.
49. Morrissey MB, Kruuk LEB, Wilson AJ. The danger of applying the breeder's equation in observational studies of natural populations. *Journal of Evolutionary Biology*. 2010; 23:2277–2288. [PubMed: 20831731]
50. Reed TE, Gienapp P, Visser ME. Testing for biases in selection on avian reproductive traits and partitioning direct and indirect selection using quantitative genetic models. *Evolution*. 2016; 70:2211–2225. [PubMed: 27464856]
51. Morrissey MB, Ferguson MM. A test for the genetic basis of natural selection: an individual-based longitudinal study in a stream-dwelling fish. *Evolution*. 2011; 65:1037–1047. [PubMed: 21108638]
52. Mills JA, et al. Archiving Primary Data: Solutions for Long-Term Studies. *Trends in Ecology & Evolution*. 2015; 30:581–589. [PubMed: 26411615]
53. Evans SR. Gauging the purported costs of public data archiving for long-term population studies. *PLoS Biology*. 2016; 14:1–9.
54. Whitlock MC, et al. A balanced data archiving policy for long-term studies. *Trends in Ecology & Evolution*. 2016; 31:84–85. [PubMed: 26708957]
55. Mills JA, et al. Solutions for archiving data in long-term studies: A reply to Whitlock et al. *Trends in Ecology & Evolution*. 2016; 31:85–87. [PubMed: 26778168]
56. Clutton-Brock T, Sheldon BC. Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*. 2010; 25:562–573. [PubMed: 20828863]
57. Gienapp P, Fior S, Guillaume F, Lasky JR, Sork VL, Csilléry K. Genomic quantitative genetics to study evolution in the wild. *Trends in Ecology & Evolution*. 2017; 32:897–908. [PubMed: 29050794]
58. Visser ME. Keeping up with a warming world; assessing the rate of adaptation to climate change. *Proceedings of the Royal Society B-Biological Sciences*. 2008; 275:649–659.
59. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine*. 2009; 6:e1000097. [PubMed: 19621072]
60. Ramakers JJC, Culina A, Visser ME, Gienapp P. Data from: Environmental coupling of heritability and selection is rare and of minor evolutionary significance in wild populations. *Dryad Digital Repository*. 2018; doi: 10.5061/dryad.3mk7f2q
61. Visser ME, et al. Effects of spring temperatures on the strength of selection on timing of reproduction in a long-distance migratory bird. *PLoS Biol*. 2015; 13:e1002120. [PubMed: 25848856]
62. Gilmour, AR., Gogel, BJ., Cullis, BR., Thompson, R. *ASReml User Guide*. Release 3.0. VSN International Ltd; 2009.
63. Butler, D., Cullis, BR., Gilmour, AR., Gogel, DJ. *ASReml-R Reference Manual*, Release 3.0. Department of Primary Industries and Fisheries; 2009.
64. Kruuk LEB. Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society B*. 2004; 359:873–890.

65. Henderson CR. Theoretical basis and computational methods for a number of different animal models. *Journal of Dairy Science*. 1988; 71:1–16.
66. Hadfield, J. MCMCglmm Course Notes. 2014. <https://cran.r-project.org/web/packages/MCMCglmm/.../CourseNotes.pdf>
67. Hadfield JD. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*. 2010; 33:1–22. [PubMed: 20808728]
68. Lillehammer M, Ødegård J, Meuwissen THE. Reducing the bias of estimates of genotype by environment interactions in random regression sire models. *Genetics Selection Evolution*. 2009; 41:30.
69. Schuurman NK, Grasman RPPP, Hamaker EL. A comparison of Inverse-Wishart prior specifications for covariance matrices in multilevel autoregressive models. *Multivariate Behavioral Research*. 2016; 51:185–206. [PubMed: 27028576]
70. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*. 2006; 1:515–534.
71. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*. 2002; 64:583–639.
72. Millar RB. Comparison of hierarchical Bayesian models for overdispersed count data using DIC and Bayes' Factors. *Biometrics*. 2009; 65:962–969. [PubMed: 19173704]
73. De Jong G. Quantitative genetics of reaction norms. *Journal of Evolutionary Biology*. 1990; 3:447–468.
74. Wood, S. Package 'mgcv': Mixed GAM Computation Vehicle with Automatic Smoothness Automation. 2017. <https://cran.r-project.org/web/packages/mgcv/mgcv.pdf>
75. Morrissey, MB., Sakrejda, K. Package 'gsg': Calculation of selection coefficients. 2015. <https://cran.r-project.org/web/packages/gsg/gsg.pdf>
76. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*. 2007; 82:591–605. [PubMed: 17944619]
77. Niemelä PT, Dingemans NJ. Meta-analysis reveals weak associations between intrinsic state and personality. *Proceedings of the Royal Society B: Biological Sciences*. 2018; 285 20172823.
78. Bates D, Bolker B, Bojesen RH, Singmann H, Grothendieck G. Package 'lme4': Linear Mixed-Effects Models using 'Eigen' and S4. CRAN. 2015
79. Nakagawa S, Santos ESA. Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology*. 2012; 26:1253–1274.
80. Falconer, DS., Mackay, TFC. Introduction to quantitative genetics. Longmans Green; 1996.
81. Reed TE, Gienapp P, Visser ME. Data from: Testing for biases in selection on avian reproductive traits and partitioning direct and indirect selection using quantitative genetic models. Dryad Digital Repository. 2016; doi: 10.5061/dryad.1255v
82. Ljungström G, Stjernstedt M, Wapstra E, Olsson M. Data from: Selection and constraints on offspring size-number trade-offs in sand lizards (*Lacerta agilis*). Dryad Digital Repository. 2016; doi: 10.5061/dryad.700jq
83. Delahaie B, Charmantier A, Chantepie S, Garant D, Porlier M, Teplitsky C. Data from: Conserved G-matrices of morphological and life-history traits among continental and island blue tit populations. Dryad Digital Repository. 2017; doi: 10.5061/dryad.03mn0
84. Teplitsky C, Mouawad NG, Balbontín J, de Lope F, Møller AP. Data from: Quantitative genetics of migration syndromes: a study of two barn swallow populations. Dryad Digital Repository. 2011; doi: 10.5061/dryad.41np7
85. Thomson CE, et al. Data from: Selection on parental performance opposes selection for larger body mass in a wild population of blue tits. Dryad Digital Repository. 2017; doi: 10.5061/dryad.m2c10
86. Wheelwright NT, Keller LF, Postma E. Data from: The effect of trait type and strength of selection on heritability and evolvability in an island bird population. Dryad Digital Repository. 2014; doi: 10.5061/dryad.dv0qt.2
87. Gienapp P, Merilä J. Data from: Disentangling plastic and genetic changes in body mass of Siberian jays. Dryad Digital Repository. 2014; doi: 10.5061/dryad.4v514

88. Mulder HA, Gienapp P, Visser ME. Data from: Genetic variation in variability: phenotypic variability of fledging weight and its evolution in a songbird population. Dryad Digital Repository. 2016; doi: 10.5061/dryad.2qv8n
89. Santure AW, et al. Data from: Replicated analysis of the genetic architecture of quantitative traits in two wild great tit populations. Dryad Digital Repository. 2015; doi: 10.5061/dryad.5t32v
90. Bonnet T, Wandeler P, Camenisch G, Postma E. Bigger is fitter? Quantitative genetic composition of selection reveals an adaptive evolutionary decline of body mass in a wild rodent population. PLoS Biology. 2017; 15:e1002592. [PubMed: 28125583]
91. Becker PJJ, Hegelbach J, Keller LF, Postma E. Data from: Phenotype-associated inbreeding biases estimates of inbreeding depression in a wild bird population. Dryad Digital Repository. 2016; doi: 10.5061/dryad.53g68
92. Evans SR, Sheldon BC. Data from: Colour in a new light: a spectral perspective on the quantitative genetics of carotenoid coloration. Dryad Digital Repository. 2015; doi: 10.5061/dryad.1cn02
93. Kim S, Fargallo JA, Vergara P, Martínez-Padilla J. Data from: Multivariate heredity of melanin-based coloration, body mass and immunity. Dryad Digital Repository. 2013; doi: 10.5061/dryad.03mn0
94. Class B, Brommer JE. Data from: A strong genetic correlation underlying a behavioural syndrome disappears during development because of genotype-age interactions. Dryad Digital Repository. 2015; doi: 10.5061/dryad.443g2
95. Becker PJJ, Hegelbach J, Keller LF, Postma E. Phenotype-associated inbreeding biases estimates of inbreeding depression in a wild bird population. Journal of Evolutionary Biology. 2016; 29:35–46. [PubMed: 26362803]
96. Delahaie B, Charmantier A, Chantepie S, Garant D, Porlier M, Teplitsky C. Conserved G-matrices of morphological and life-history traits among continental and island blue tit populations. Heredity. 2017; 119:76–87. [PubMed: 28402327]
97. Thomson CE, et al. Selection on parental performance opposes selection for larger body mass in a wild population of blue tits. Evolution. 2017; 71:716–732. [PubMed: 28106259]
98. Kim S, Fargallo JA, Vergara P, Martínez-Padilla J. Multivariate heredity of melanin-based coloration, body mass and immunity. Heredity. 2013; 111:139–146. [PubMed: 23591519]
99. Teplitsky C, Mouawad NG, Balbontín J, de Lope F, Møller AP. Quantitative genetics of migration syndromes: a study of two barn swallow populations. Journal of Evolutionary Biology. 2011; 24:2025–2039. [PubMed: 21707815]
100. Ljungström G, Stjernstedt M, Wapstra E, Olsson M. Selection and constraints on offspring size-number trade-offs in sand lizards (*Lacerta agilis*). Journal of Evolutionary Biology. 2016; 29:979–990. [PubMed: 26851437]
101. Evans SR, Sheldon BC. Colour in a new light: a spectral perspective on the quantitative genetics of carotenoid coloration. Functional Ecology. 2015; 29:96–103.
102. Santure AW, et al. Replicated analysis of the genetic architecture of quantitative traits in two wild great tit populations. Molecular Ecology. 2015; 24:6148–6162. [PubMed: 26661500]
103. Mulder HA, Gienapp P, Visser ME. Phenotypic variability of fledging weight and its evolution in a songbird: do families differ genetically? Evolution. 2016; 70:2004–2016. [PubMed: 27435892]
104. Gienapp P, Merilä J. Disentangling plastic and genetic changes in body mass of Siberian jays. Journal of Evolutionary Biology. 2014; 27:1849–1858. [PubMed: 24930732]
105. Wheelwright NT, Keller LF, Postma E. The effect of trait type and strength of selection on heritability and evolvability in an island bird population. Evolution. 2014; 68:3325–3336. [PubMed: 25130048]

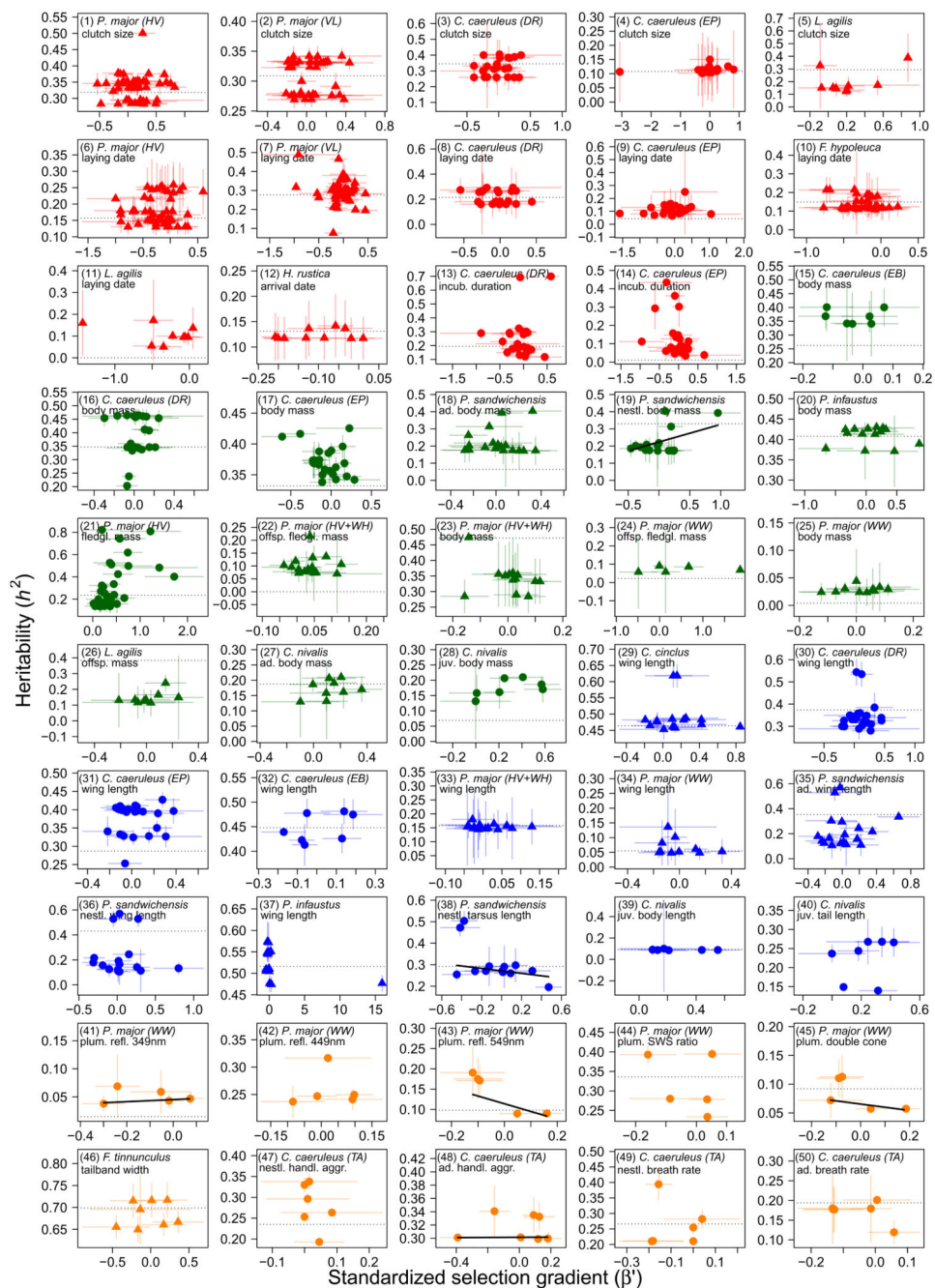


Figure 1. Heritability as a function of the standardized selection gradient.

Standard errors (SEs) are omitted when $SE_{h^2} > 0.5$ and $SE_{\beta'} > 1$ for visual aid. Regression lines result from weighted least-squares regression models (weights: $1/[(SE_{h^2})^2]$), with bootstrapping to account for uncertainty in β' , shown only when the 95% CI did not include zero. Colours denote different trait classes (red: life history; green: body mass; blue: morphology; orange: miscellaneous), whereas shapes indicate selection based on survival (circles) or based on number of fledglings or recruits (triangles). Dotted horizontal lines denote the constant heritability as estimated from a standard animal model. Duplicate traits

(from same population but different dataset) are not shown. Data sources by panel: **(1,2,6,7)** ref. 81; **(5,11,26)** ref. 82; **(3,4,8,9,13,14,16,17,30,31)** ref. 83; **(10)** unpubl. data; **(12)** ref. 84; **(15,32)** ref. 85; **(18,19,35,36,38)** ref. 86; **(20,37)** ref. 87; **(21)** ref. 88; **(22–25,33,34)** ref. 89; **(27,28,39,40)** ref. 90; **(29)** ref. 91; **(41–45)** ref. 92; **(46)** ref. 93; **(47–50)** ref. 94.

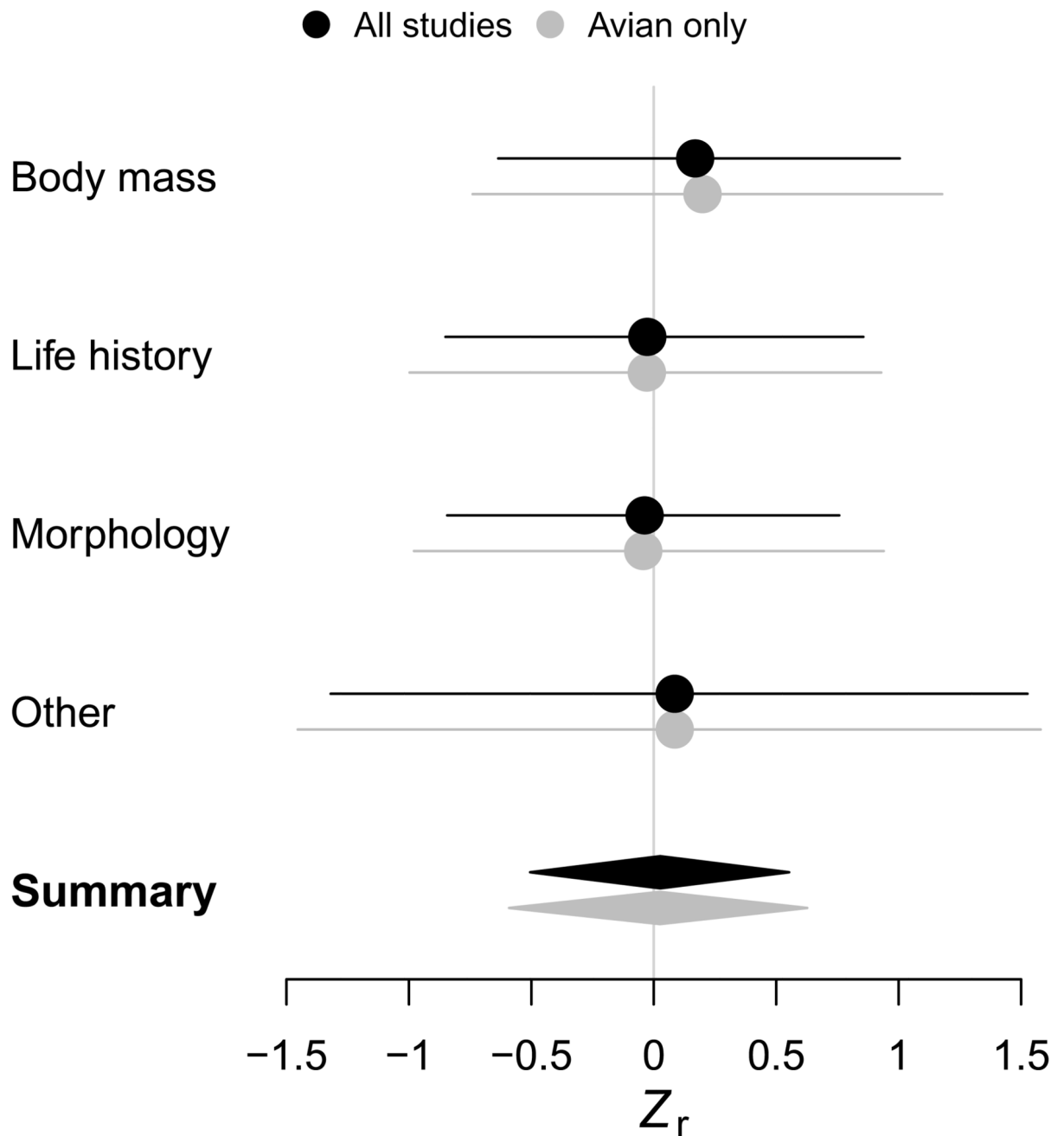


Figure 2. Meta-analysis on the heritability–selection correlation coefficients.

Coefficients r were standardised using Fisher’s Z transformation prior to analysis. Estimates and bootstrapped 95% CIs are shown, predicted from a linear mixed-effects model and unconditioned on the random term ‘study area’. The summary statistic results from a model that included only the intercept as a fixed term. Estimates from an analysis excluding non-avian traits are shown for comparison.

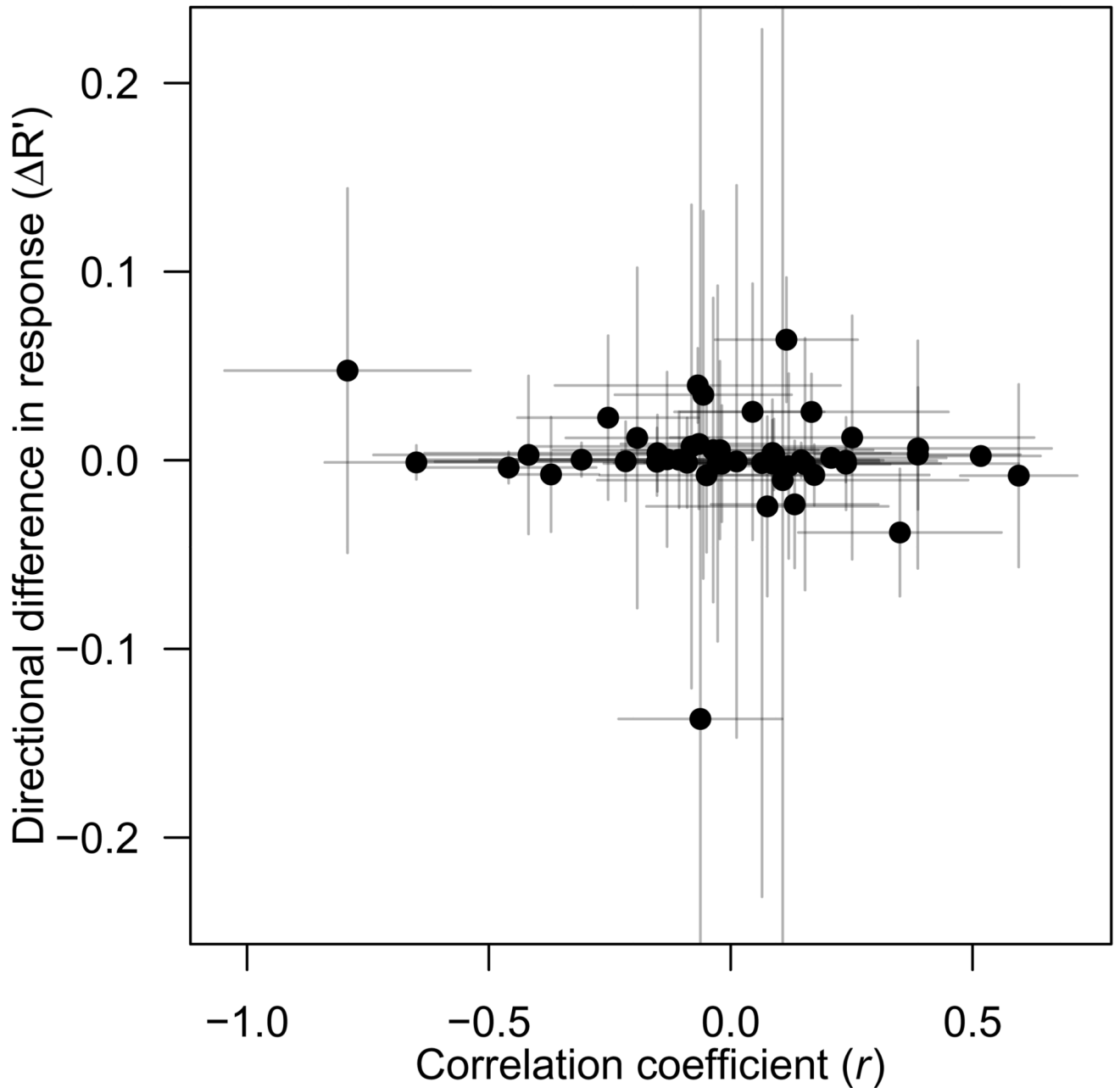


Figure 3. No effect of a correlation between heritability and selection on differences in selection response.

Correlation coefficients ($r \pm$ standard errors) result from WLS regressions of heritability against standardised selection gradients; R' (\pm standard errors) is the mean, directional difference between expected responses to selection assuming varying vs. constant heritability. Each data point represents a single trait–species–population combination.

Table 1
Overview of studies used in the gene-by-environment and selection analyses.

The studies and traits listed are the ones that met inclusion criteria (see Supplementary Table S1 for an untrimmed overview of datasets) and could be successfully analysed.

Species	Data source (refs.)	Locality	Study aim of the original paper	Trait	N _{obs}	N _{ind}	N _{year}	<i>h</i> ² (s.e.)	Selection used in analysis
<i>Cinclus cinclus</i>	91, 95	ZU	Estimate biases in inbreeding depression	Wing length (mm)	1132	672	18	0.464 (0.066)	F
<i>Chyonomis nivalis</i>	90	CW	Predict genetic changes in body mass	Body mass (g)	3382	931	9	a: 0.188 (0.064)	F
				Body length (mm)*	2761	791	8	j: 0.069 (0.056)	V
				Tail length (mm)*	3382	931	9	j: 0.093 (0.070)	V
								a: 0.281 (0.079)	–
								j: 0.101 (0.054)	V
<i>Cyanistes caeruleus</i>	41, 94	TA	Estimate the developmental stability of behavioural syndromes	Nestling handling aggression	6149	6149	8	0.235 (0.043)	V
				Adult handling aggression	1633	1103	8	0.283 (0.057)	V
				Nestling breath rate (breaths/s)	5863	5863	7	0.266 (0.037)	V
				Adult breath rate (breaths/s)	1526	1031	7	0.194 (0.063)	V
	83, 96	DR	Investigate spatial variation in G-matrix in populations with contrasting population history and selective environment	Incubation duration (d)	1104	740	24	0.195 (0.074)	V
				Laying date	1104	740	24	0.214 (0.094)	V
				Clutch size	1104	740	24	0.345 (0.104)	V
				Wing length (mm)	2916	1597	24	0.374 (0.040)	V
				Body mass (g)	2916	1597	24	0.347 (0.038)	V
		EP	As above	Incubation duration (d)	997	637	35	0.011 (0.102)	V
				Laying date	997	637	35	0.043 (0.148)	V
				Clutch size	997	637	35	0.108 (0.145)	V
				Wing length (mm)	2260	1187	26	0.287 (0.065)	V
				Body mass (g)	2260	1187	26	0.332 (0.065)	V

Species	Data source (refs.)	Locality	Study aim of the original paper	Trait	N _{obs}	N _{ind}	N _{year}	<i>t</i> ² (s.e.)	Selection used in analysis
	85, 97	EB	Quantify selection on parental care to explain stasis in evolution of offspring body size	Wing length (mm)	1677	847	8	0.448 (0.151)	V
<i>Falco tinnunculus</i>	93, 98	SP	Quantify multivariate heredity of colouration, mass and immunity	Body mass (g) Tail-band width (mm)	1677 688	847 444	8 17	0.262 (0.117) 0.699 (0.086)	V F
<i>Ficedula hypoleuca</i>	60	HV	None (published for the purpose of this paper)	Laying date	3044	2211	39	0.149 (0.072)	F
<i>Hirundo rustica</i>	84, 99	SP	Estimate genetic correlation between arrival date and life-history traits	Spring arrival date	2337	1407	17	0.131 (0.076)	F
<i>Lacerta agilis</i>	82, 100	AS	Test for trade-off between offspring size and number	Clutch size	472	288	10	0.294 (0.066)	F
<i>Parus major</i>	92, 101	WW	Estimate genetic variance in colour expression across the visual spectrum	Laying date Mean offspring mass (g) [‡] Plumage reflectance at 349 nm [‡]	370 452 2904	236 279 1618	9 10 6	0 0.384 (0.064) 0.015 (0.035)	F F V
				Plumage reflectance at 449 nm [‡]	2901	1616	6	0.194 (0.051)	V
				Plumage reflectance at 549 nm [‡]	2901	1616	6	0.098 (0.043)	V
				SWS ratio (plumage reflectance)	2901	1616	6	0.336 (0.051)	V
				Double cone (plumage reflectance)	2901	1616	6	0.092 (0.044)	V
				Wing length (mm)	2892	1614	6	0.478 (0.054)	V
				Body mass (g)	2878	1613	6	0.357 (0.052)	V
	89, 102	WW	Investigate the genetic architect of a suite of parameters in two populations	Adult body mass	2919	1358	12	0.004 (0.034)	V
				Offspring fledgling weight (g) [‡]	3162	328	13	0.022 (0.107)	V
				Wing length	3206	1408	12	0.055 (0.042)	V
				Adult body mass (g)	1543	477	16	0.472 (0.027)	F [§]
				Clutch size	1585	943	17	0.058 (0.181)	F [§]
				Offspring fledgling weight (g) [‡]	8569	744	17	0	F [§]
				Wing length	1908	1275	17	0.158 (0.140)	F [§]

Species	Data source (refs.)	Locality	Study aim of the original paper	Trait	N _{obs}	N _{ind}	N _{year}	<i>h</i> ² (s.e.)	Selection used in analysis
	50, 81	HV	Test for bias in selection on life-history traits	Clutch size	4054	2861	57	0.318 (0.055)	F
		VL	As above	Laying date	4054	2861	57	0.157 (0.052)	F
				Clutch size	3700	2368	52	0.306 (0.044)	F
				Laying date	3700	2368	52	0.277 (0.048)	F
	88, 103	HV	Estimate heritability of within-family variance in fledgling weight	Fledgling weight (g)	17535	17535	36	0.235 (0.027)	V
<i>Perisoreus infaustus</i>	87, 104	OB	Disentangle plastic and genetic changes in body mass	Clutch size	2175	1598	36	0.282 (0.082)	F
				Body mass (g)	1619	1025	30	0.408 (0.058)	F
<i>Passerculus sandwichensis</i>	86, 105	KI	Investigate the relationship between heritability/evolvability and selection	Wing length (mm)	1453	1016	28	0.516 (0.056)	F
				Day-8 to Yr-1 wing length (mm)	2839	2469	20	a: 0.353 (0.072)	F
				Day-8 to Yr-1 tarsus length (mm)*	1913	1615	20	a: 0.398 (0.070)	–
				Day-8 to Yr-1 body mass (g)	2469	2362	20	a: 0.064 (0.018)	F
								j: 0.330 (0.059)	V

Note. Locality: AS = Asketunnan, Sweden; CW = Churwalden, Switzerland; DR = D-Rouvriere, France; EB = Edinburgh, UK; EP = E-Prrio, France; HV = Hoge Veluwe, NL; KI = Kent Island, Canada; OB = Ostrobothnia, Finland; SP = Spain; TA = Tammisaari, Finland; VL = Vieland, the Netherlands; WH = Westerheide, NL; WW = Wytham Woods / Bagley Woods, UK; ZU = Zürich, Switzerland.

Selection: F = fecundity; V = viability; ‘-’ = disregarded due to fixed nature of trait in adults.

Nobs/Nind/Nyear: number of observations/individuals/years (environments); *h*²: narrow-sense trait heritability (a: adult; j: juvenile; significant values in boldface).

* ‘Fixed’ trait: may change from juvenile to adult stage but are assumed to be relatively constant within adult lifespan.

[†]Trait considered maternal.

[‡]Trait constitutes one out of a range of 198 2-nm bands; three bands equally spaced apart and spanning most of the gradient were chosen for analysis.

[§]Fecundity based on number of fledglings, not recruits.

Table 2
Predicted selection response assuming constant vs. environment-dependent heritability.

Predicted response (R') differed in absolute terms from year to year under the two approaches for all six cases where a correlation between heritability and selection was found; in none of these cases the difference was in a consistent direction.

Species	Trait	$R'_{absolute}$ [95% CI]	$R'_{directional}$ [95% CI]
<i>Cyanistes caeruleus</i>	Adult handling aggression	0.031 [0.015, 0.049]	0.005 [-0.026, 0.031]
<i>Parus major</i>	Plumage refl. (at 349 nm)	0.007 [0.003, 0.012]	0.005 [-0.001, 0.011]
	Plumage refl. (at 549 nm)	0.012 [0.005, 0.021]	0.005 [-0.007, 0.016]
	Double cone plumage refl.	0.008 [0.003, 0.014]	-0.001 [-0.010, 0.007]
<i>Passerculus sandwichensis</i>	Nestling tarsus length	0.059 [0.038, 0.085]	0.005 [-0.034, 0.046]
	Nestling body mass	0.072 [0.046, 0.101]	-0.019 [-0.062, 0.024]

Note. R' is measured in phenotypic standard deviations. Estimates of differences were calculated using bootstrapping procedures