Testing for genetic assimilation with phylogenetic comparative analysis: Conceptual, methodological, and statistical considerations

Alex R. Gunderson[1,2] iD and Liam J. Revell[3,4] iD

[1]Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, Louisiana 70118

[2]E-mail: agunderson@tulane.edu

[3]Department of Biology, University of Massachusetts Boston, Boston, Massachusetts 02125

[4]Facultad de Ciencias, Universidad Católica de la Santísima Concepción, Concepción, Chile

Genetic assimilation is a process that leads to reduced phenotypic plasticity during adaptation to novel conditions, a potentially important phenomenon under global environmental change. Null expectations when testing for genetic assimilation, however, are not always clear. For instance, the statistical artifact of regression to the mean could bias us toward detecting genetic assimilation when it has not occurred. Likewise, the specific mechanism underlying plasticity expression may affect null expectations under neutral evolution. We used macroevolutionary numerical simulations to examine both of these important issues and their interaction, varying whether plasticity evolves, the evolutionary mechanism, trait measurement error, and experimental design. We also modified an existing reaction norm correction method to account for phylogenetic nonindependence. We found (1) regression to the mean is pervasive and can generate spurious support for genetic assimilation; (2) experimental design and post hoc correction can minimize this spurious effect; and (3) neutral evolution can produce patterns consistent with genetic assimilation without constraint or selection, depending on the mechanism of plasticity expression. Additionally, we reanalyzed published macroevolutionary data supporting genetic assimilation, and found that support was reduced after proper correction. Considerable caution is thus required whenever investigating genetic assimilation and reaction norm evolution at macroevolutionary scales.

KEY WORDS: Acclimation, evolution, global change, plasticity, reaction norm.

Phenotypic plasticity is an important phenomenon in evolutionary biology with far-reaching implications (Ghalambor et al. 2007; Murren et al. 2015). For example, plasticity in a given trait can facilitate evolutionary radiations (Pfennig et al. 2010), contribute to local adaptation (Price et al. 2003), and dampen selection on underlying genetic variation (Price et al. 2003; Crispo 2008). Interest in the expression and evolution of phenotypic plasticity has increased in recent years due to the intense selection pressure that human activity has imposed on the natural world (Chevin et al. 2010; Kingsolver and Buckley 2017). Taxa with greater plasticity in traits that are important for tolerating human-induced changes to the environment are predicted to fare better under the novel conditions created by global warming, urbanization, and species introductions (Somero 2010; Gunderson and Stillman 2015; Seebacher et al. 2015). For instance, plasticity in thermal physiology is predicted to decrease the negative impact of global warming on ectotherms (Gunderson et al. 2017; Riddell et al. 2018; Rohr et al. 2018; Morley et al. 2019). Greater knowledge of the forces that shape and constrain phenotypic plasticity is crucial both for our fundamental understanding of ecology and evolution and our ability to predict and mitigate the consequences of global change.

Adaptation to novel conditions is sometimes associated with a reduction or loss of plasticity in traits under selection, a

process that is known as genetic assimilation (Waddington 1953; Schlichting and Pigliucci 1998). Some have argued that genetic assimilation is a key mechanism of adaptive evolutionary change (West-Eberhard 2003; Crispo 2007). For example, the canalization of formerly induced phenotypes can provide a means of adaptive phenotypic divergence during local adaptation and adaptive radiation (Ehrenreich and Pfennig 2016; Martin et al. 2016; Gunter et al. 2017; Schneider and Meyer 2017). In addition, genetic assimilation may be an outcome of population responses to anthropogenic global change. Lande (2009) found that selection due to dramatic changes in the environment should result in the evolution of increased phenotypic plasticity followed by genetic assimilation.

The intersection of adaptation, genetic assimilation, and global change is highlighted by the Baseline Tolerance/Tolerance Plasticity Trade-off Hypothesis (hereafter referred to as the Trade-off Hypothesis). Emerging from the field of evolutionary physiology, the hypothesis states that, as organisms evolve greater baseline tolerance to extreme temperatures (heat or cold), plasticity in thermal tolerance will decrease (van Heerwaarden and Kellermann 2020). In other words, high levels of constitutive thermal tolerance evolve by genetic assimilation, resulting in thermal tolerance phenotypes that are less sensitive to environmental temperature variation (Sikkink et al. 2014). Support for this hypothesis comes from a variety of organisms, including terrestrial and aquatic vertebrates and invertebrates, at both the intraspecific and interspecific levels (e.g., Stillman 2003; Esperk et al. 2016; Comte and Olden 2017; Armstrong et al. 2019). One of several implications of the Trade-off Hypothesis is that selection to evolve greater baseline heat tolerance under global change will result in reduced capacity to track subsequent warming via plasticity (van Heerwaarden and Kellermann 2020).

Hypotheses that make predictions about trait change over time among individuals and/or taxa, such as the Trade-off Hypothesis, are prone to spurious support due to the well-known statistical phenomenon called *regression to the mean*. A consequence of this phenomenon is that when a particular trait is measured repeatedly, the experimental units (e.g., species, populations, or individuals) that start with particularly high or low values relative to the population mean are also likely to record the greatest change in value over time as subsequent measures regress toward the mean (Barnett et al. 2005). This is not a real biological effect: rather, it is an artifact created by sampling and measurement error. Regression to the mean is a common problem in repeated-measures analyses, and statistical methods have been developed to overcome the spurious results that it can create. For example, data resampling techniques have been applied to generate null distributions of trait change (Jackson and Somers 1991), and methods to statistically remove regression to the mean effects prior to statistical analysis have also been pro-

posed (Kelly and Price 2005). These approaches are most often employed at the intraspecific level (Ghalambor et al. 2015; Deery et al. 2021; O'Neill et al. 2021), but accounting for regression to the mean is also necessary, although more often neglected, in interspecific macroevolutionary analyses (Baker et al. 2015).

Motivated by current debate surrounding the Trade-off Hypothesis and increasing interest in the evolution of phenotypic plasticity and genetic assimilation under global change (Gunderson and Stillman 2015; Kelly 2019; Sasaki and Dam 2019; van Heerwaarden and Kellermann 2020; Barley et al. 2021; Sasaki and Dam 2021; Preston et al. 2022), we used simulations to explore the more general consequences of regression to the mean for tests of genetic assimilation at the macroevolutionary level. We conducted simulations that assumed mechanistically different models for plasticity evolution, as well as different experimental sampling designs for measuring species phenotypic traits. In addition, we conducted simulations with and without the assumption that phenotypes are measured with error. We found that measurement error alone can be sufficient to recover a macroevolutionary pattern that is consistent with genetic assimilation due entirely to the phenomenon of regression to the mean. This problem can be avoided, however, by using an appropriate experimental design. In addition, we found that macroevolutionary patterns consistent with genetic assimilation are the null expectation under random Brownian motion evolution for some models by which plasticity evolves. Finally, we applied a statistical method designed to remove the regression to the mean effect from compromised data (Kelly and Price 2005) that we modified to take nonindependence due to phylogeny into account. We found that our approach can reduce the effect of regression to the mean in some, but not all, circumstances. We applied this modified approach to published phylogenetic comparative data that supported the Trade-off Hypothesis and showed that the support decreased with proper data correction.

Methods

In our simulations, we assumed that there were two different environments, baseline (ancestral) and novel, that induce different phenotypes. The difference between trait values in each environment represented the magnitude of phenotypic plasticity and, in our two-environment system, is equivalent to the slope of the reaction norm. In any given simulation, we first generated a random phylogenetic tree with 50 terminal nodes using the *pbtree* function in the phytools R package (Revell 2012; R Development Core Team 2021). Depending on the simulation, we either modeled the evolution of plasticity across the phylogeny using the *fastBM* function or assumed no phenotypic evolution at all (see below for details). In addition, we assumed a certain degree of
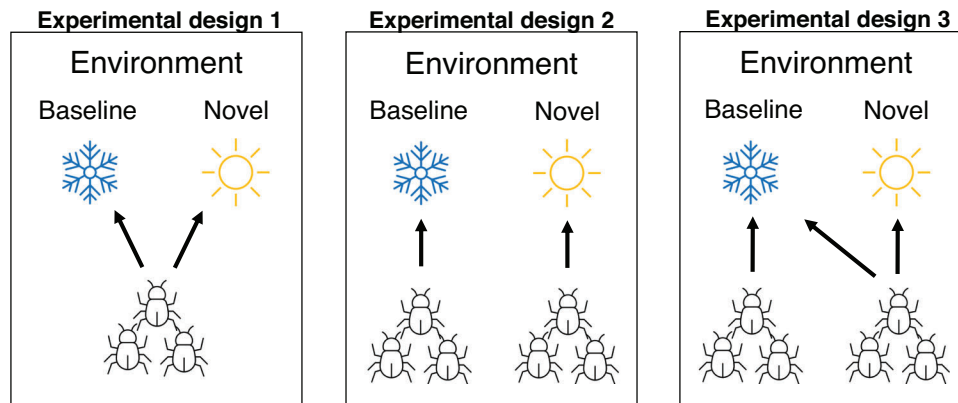
Figure 1. Three experimental designs that we simulated for estimating baseline phenotype and phenotypic plasticity of species. In Experimental Design 1, baseline phenotype and plasticity are estimated from a single group of individuals that have their phenotype measured twice: once after exposure to each environment. In Experimental Design 2, there are two groups of individuals that have their phenotypes measured only once, after exposure to either the baseline (ancestral) or novel environment. In Experimental Design 3, there are also two groups of individuals. One group of individuals has its phenotypes measured only after exposure to the baseline condition and they are used only to estimate the baseline phenotype. The other group of individuals has its phenotypes measured twice (as in Experimental Design 1), but they are only used to estimate phenotypic plasticity. Thermal environments are shown simply as an example.

error in the measurement of population mean phenotypes in each environment. We did so by drawing phenotypes for individuals randomly from a Gaussian distribution with a mean value at the true phenotypic population mean. In practice, this type of error can emerge both from random sampling of individuals within a variable population and from methodological or instrument imprecision. Because both types of error can contribute to the phenomenon of regression to the mean, we effectively combined them in our simulations and collectively refer to them as "measurement error." For each simulation run, we calculated the phylogenetic correlation coefficient for the association between the baseline phenotype and the reaction norm slope using the *phyl.vcv* function of phytools. Starting conditions assumed a baseline phenotypic trait value of 0 and a phenotypic value in the novel environment of 1. In simulations where evolution occurred, the Brownian motion rate parameter (σ^2) was set at 1, 0.1, or 0.01 for respective traits. Phenotypic measurement error was set to have variances (s^2) of 1, 0.1, or 0.01.

We predicted that experimental design would influence whether regression to the mean biased our results. As such, we modeled three different experimental designs for measuring the phenotypes of each taxon in each environment (Fig. 1). In Design 1, we assumed that each individual has their phenotype measured twice in total: once after acclimation to each environment. Therefore, in this case the same individuals are used to estimate both the baseline phenotype and the reaction norm. In Design 2, we assumed that different individuals are acclimated to each environment prior to phenotypic measurement. In this case, individuals exposed to the baseline environment are used to estimate the baseline phenotype, and the difference in phenotype between

the two sets of individuals is used to estimate the reaction norm. Finally, in Design 3, we assumed that one set of individuals is acclimated only to the baseline environment prior to phenotypic trait measurement, and that a second set of individuals have their phenotypes measured twice: once after acclimation to each of the two environments. The former individuals are used only to estimate baseline phenotypes, and the latter individuals are used only to estimate the slope of the reaction norm.

We tested whether regression to the mean could emerge as a significant problem when phenotypic plasticity evolves assuming a neutral evolutionary process (in this case, Brownian motion). We used two different models for the evolution of phenotypic plasticity, that we refer to as *Linked Phenotypes* and *Unlinked Phenotypes*. In the Linked Phenotypes model, we assumed that the slope and intercept of the reaction norms evolve directly, with the phenotype that is expressed in each environment evolving indirectly as a by-product. This is one of the most common ways in which the evolution of phenotypic plasticity is conceptualized and studied theoretically (Lande 2009; Chevin et al. 2010, 2013). We refer to this as the Linked Phenotypes model because the phenotypes that are expressed in each environment are each determined by the slope and the intercept, linking them under a neutral evolutionary process. In the Unlinked Phenotypes model, the phenotype that is expressed in a given environment evolves directly and independently of the phenotype expressed in the other environment. Therefore, the phenotypes that are expressed in each environment lack genetic covariance (Via and Lande 1985; Ghalambor et al. 2007), and the slope and intercept of the reaction norm emerge merely as a by-product of the trait values expressed in each environment (Via 1993; De Jong

1995). Using each model, we conducted simulations in which the evolving parameters diverge via Brownian motion, and then estimated the phylogenetic correlation coefficient for an association between baseline phenotype and the reaction norm slope, as described above.

When possible, we applied a statistical correction developed by Kelly and Price (2005) to remove the regression to the mean effect from reaction norm values. The method was not originally designed for use with phylogenetic comparative data, but we modified it to that purpose here. We use the correction only with data simulated under Experimental Designs 1 and 2, as the method cannot be applied to Experimental Design 3. The approach is based on the premise that if the plastic response is constrained by the baseline phenotype, then phenotypic variation should be reduced under novel conditions and that phenotypes in the baseline and novel conditions should be correlated within experimental units (Kelly and Price 2005). The approach is implemented as follows. First, we estimate the phylogenetic variance in phenotypes in each environment (s_i) using the *phyl.vcv* function, then we conduct a Pittman's test of homogeneity of variance. If there is no difference in variance between the phenotypes, then corrected reaction norms ($\hat{D}^*$) are calculated using the following function:

$$\hat{D}^* = r\left(X_1 - \bar{X}_1\right) - \left(X_2 - \bar{X}_2\right), \tag{1}$$

where r is the phylogenetic correlation coefficient for the relationship between phenotypes in the baseline and novel environments, X_i is the vector of phenotypes in a given environment, and $\bar{X}_i$ is the phylogenetic phenotypic mean in a given environment. If the phenotypic variance differs between environments, then an adjusted correlation coefficient ($\hat{r}$) is used in equation (1), calculated as follows:

$$\hat{r} = \frac{2rs_1s_2}{s_1^2 + s_2^2}, \tag{2}$$

where s_i^2 is the estimated phylogenetic variance in the environment i. $\hat{D}^*$ is then used as the reaction norm value for subsequent analyses as described above.

Using our modified Kelly and Price (2005) approach, we also reanalyzed previously published data on heat tolerance plasticity across Nudibranch species (Armstrong et al. 2019). The authors of that study used what we refer to as Experimental Design 2 (Fig. 1), and originally found support for the Trade-off Hypothesis, as Nudibranch species with greater baseline heat tolerance had reduced heat tolerance plasticity. However, they did not adjust values for potential regression to the mean as we do here.

Results
RESULTS WITHOUT EVOLUTION BUT WITH MEASUREMENT ERROR

We first present results of simulations in which phenotypic plasticity is present but does not diverge among taxa. In this case, the only source of variation among taxa is measurement error. If regression to the mean is not an issue, we would expect the correlation coefficients generated in our simulations to be centered on 0.

Instead, we found that measurement error generated a pattern consistent with genetic assimilation (i.e., a negative correlation between baseline phenotype and reaction norm slope) due to regression to the mean, but only for Experimental Designs 1 and 2 (Fig. 2). By contrast, regression to the mean does not bias results using Experimental Design 3 (Fig. 2). This finding does not depend on the level of plasticity organisms exhibit. For example, even if there is no plasticity in the trait, a negative relationship between baseline phenotype and reaction norm slope will emerge with Experimental Designs 1 and 2 if the phenotypes in each environment are measured with error (see Fig. S1; see also Deery et al. 2021 for an individual-level example). The bias induced by Experimental Designs 1 and 2 is largely removed, however, when the simulated data are adjusted using our phylogenetic modification of the Kelly and Price (2005) correction (Fig. 2, "corrected estimates").

RESULTS WITH NEUTRAL EVOLUTION OF PHENOTYPIC PLASTICITY

A pattern consistent with genetic assimilation is unlikely to evolve under neutral processes if phenotypic plasticity evolves via the Linked Phenotype mechanism (i.e., if reaction norm slope and intercept evolve directly; Fig. 3, "true" values). However, this assumes that there is no measurement error when estimating phenotypes. When measurement error is considered, regression to the mean can produce spurious support for genetic assimilation with Experimental Designs 1 and 2, although this effect depends on the magnitude of measurement error relative to the rate of evolution (Fig. 3, "estimates" values). The greater the measurement error, the more likely that regression to the mean becomes a problem. Adjusting data generated by Experimental Designs 1 and 2 using our modified Kelly and Price correction removes the regression to the mean effect (Fig. 3, "corrected estimates"). These results, however, also highlight the importance of judicious application of the Kelly and Price correction. This is because if there is a weakly negative to positive correlation between baseline phenotype and plasticity, the Kelly and Price correction can *create* a spurious negative relationship between the two variables (e.g., see Fig. 3 results with Experimental Design 2, $\sigma^2 = 1$ and $s^2 = 0.01$).
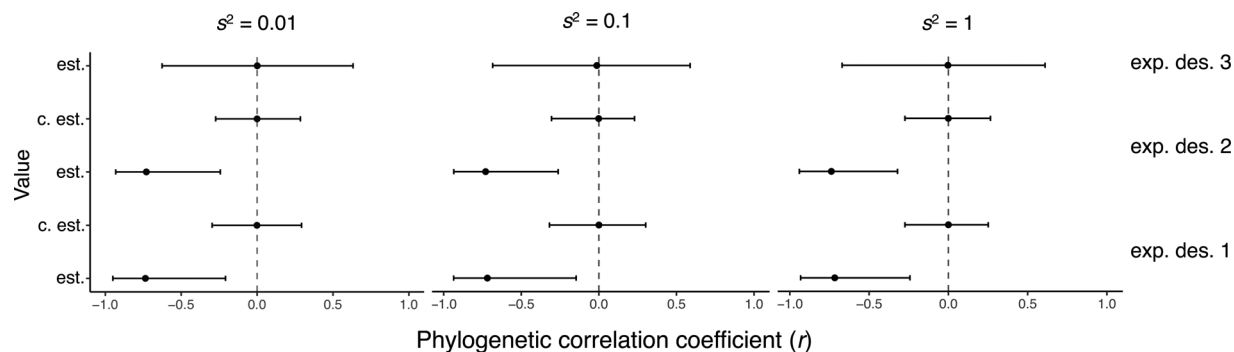
Figure 2. Summary of phylogenetic correlation coefficients between baseline phenotype and reaction norm slope from simulations in which plasticity is present but does not diverge among taxa. Phenotypes in each environment are measured with error. Points indicate the median and black lines the central 95% of correlation coefficients from 1000 simulations. The error variance used in simulations is shown above each column. "est." denotes estimated correlation coefficients given measurement error, whereas "c. est" denotes those same correlation coefficients after data adjustment to remove the effect of regression to the mean following our modified Kelly and Price (2005) correction. Corrected results are not given for Experimental Design 3 because the method cannot be applied to data collected with that design.

If phenotypic plasticity evolves via the Unlinked Phenotypes mechanism (phenotypes in each environment evolve directly and independently of one another), neutral evolution will yield a negative relationship between baseline phenotype and reaction norm slope (Fig. 4, "true" values). Furthermore, the negative relationship between these traits is maintained in the presence of measurement error (Fig. 4, "estimated" values). In this case, however, adjusting the data using the Kelly and Price correction mostly eliminates the true negative relationships between the traits (Fig. 4, "corrected" estimates). Conversely, if data are collected using Experimental Design 3, the true negative relationship between the traits is largely maintained. Nonetheless, as measurement error is increased the capacity for Experimental Design 3 to yield the true negative relationship decreases (Fig. 4).

REANALYSIS OF NUDIBRANCH HEAT TOLERANCE DATA

We reanalyzed data on heat tolerance plasticity in Nudibranchs (Armstrong et al. 2019) by applying our phylogenetic adaptation of the Kelly and Price correction. In the original analysis, the authors found a significant negative relationship between independent contrasts for baseline heat tolerance and heat tolerance plasticity ($P < 0.005$; Armstrong et al. 2019). In our reanalysis, we found that a negative relationship remained, but was no longer significant at the $\alpha = 0.05$ level ($P = 0.069$; Fig. 5). In addition, the variance explained dropped from 88% to 51%. The maintenance of a negative relationship and relatively high explanatory power suggests that there may be a real biological relationship between baseline tolerance and plasticity in this group, but the need to adjust for regression to the mean raises the bar for the evidence required.

Discussion

Herein, we have identified three general patterns important for testing and understanding the evolution of genetic assimilation at the macroevolutionary level: (1) regression to the mean is a pervasive problem; (2) experimental design and data correction can minimize the effect of regression to the mean; and (3) neutral evolutionary processes can produce patterns consistent with genetic assimilation, depending on the linkage between trait expression in different environments. Below, we discuss each of these findings in greater detail.

REGRESSION TO THE MEAN WHEN PLASTICITY DOES NOT DIVERGE

When phenotypic plasticity is present but does not diverge among taxa, measurement error alone generates spurious correlations between baseline phenotype and plasticity under Experimental Designs 1 and 2 (Fig. 2, "estimated" values). Furthermore, the magnitude of the correlation coefficient was affected very little by the magnitude of simulated measurement error (Fig. 2). This is because error variation cannot be swamped by biological variation when there are no phenotypic differences among taxa. These results clearly demonstrate that regression to the mean can be a significant problem when testing for genetic assimilation at the macroevolutionary level. However, the choice of experimental design matters considerably. When data are collected using Experimental Design 3, regression to the mean is not a persistent problem (Fig. 2).

Why are Experimental Designs 1 and 2 susceptible to regression to the mean while Experimental Design 3 is not? The crux of the matter is that with Designs 1 and 2, there is only one estimate of baseline phenotype. That single baseline pheno-
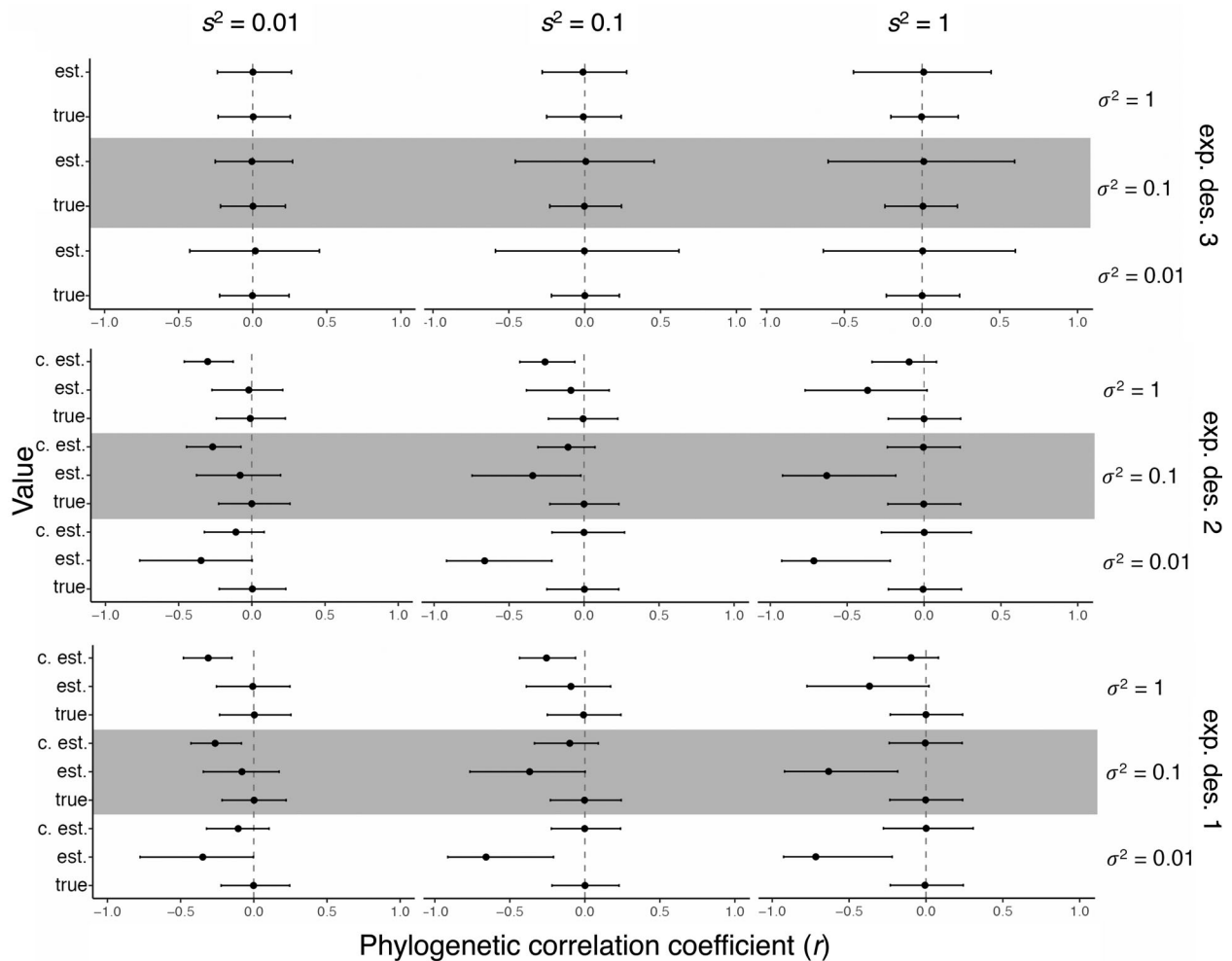
Figure 3. Summary of phylogenetic correlation coefficients between baseline phenotype and reaction norm slope from simulations in which plasticity evolved by Brownian motion via the Linked Phenotypes mechanism. Each panel shows the result of 1000 simulations. Each row shows the results of simulations with a different experimental design (Fig. 1): *Bottom*: Design 1. *Middle*: Design 2. *Top*: Design 3. Alternating shaded areas differentiate simulations with different Brownian motion rate parameters (σ^2). Each column contains simulations with a different error variance for phenotypic estimation (s^2). "true" denotes the true phylogenetic correlation values from the simulations. "est." denotes estimates of the true correlation coefficients given measurement error. "c. est" denotes estimated correlation coefficients after data adjustment to remove the effect of regression to the mean following our modified Kelly and Price (2005) correction. Corrected results are not given for Experimental Design 3 because the method cannot be applied to data collected with that design.

type estimate is also used to estimate plasticity. If measurement error causes an estimate of baseline phenotype to be above the true value, that will drive down the estimated reaction norm slope. Conversely, if error causes a baseline estimate to be below its true value, this will in turn drive up the estimated reaction norm slope.

Design 3 demands two independent estimates of the phenotype expressed under baseline conditions. One of those estimates (let us call it e1) is used only as the baseline phenotype estimate but does not contribute to the calculation of plasticity. The other estimate (e2) contributes only to the calculation of plasticity. Under this design, if measurement error causes e1 to be above or below the true value it will have no effect on the plasticity estimate, which depends on e2.

All hope is not lost with Experimental Designs 1 and 2, however. Using our modified Kelly and Price correction, phenotypic values can be adjusted such that the bias toward a spurious negative relationship is removed (Fig. 2, "corrected" values; although see below for caveats). Next, we discuss the implications of these different approaches when plasticity itself evolves.

RESULTS UNDER NEUTRAL EVOLUTION OF PLASTICITY: LINKED PHENOTYPE MODEL

Under the Linked Phenotypes model, the intercept (essentially, the baseline phenotype) and reaction norm slope evolve via Brownian motion. Under this model, neutral evolution does not produce a negative relationship between baseline phenotype and
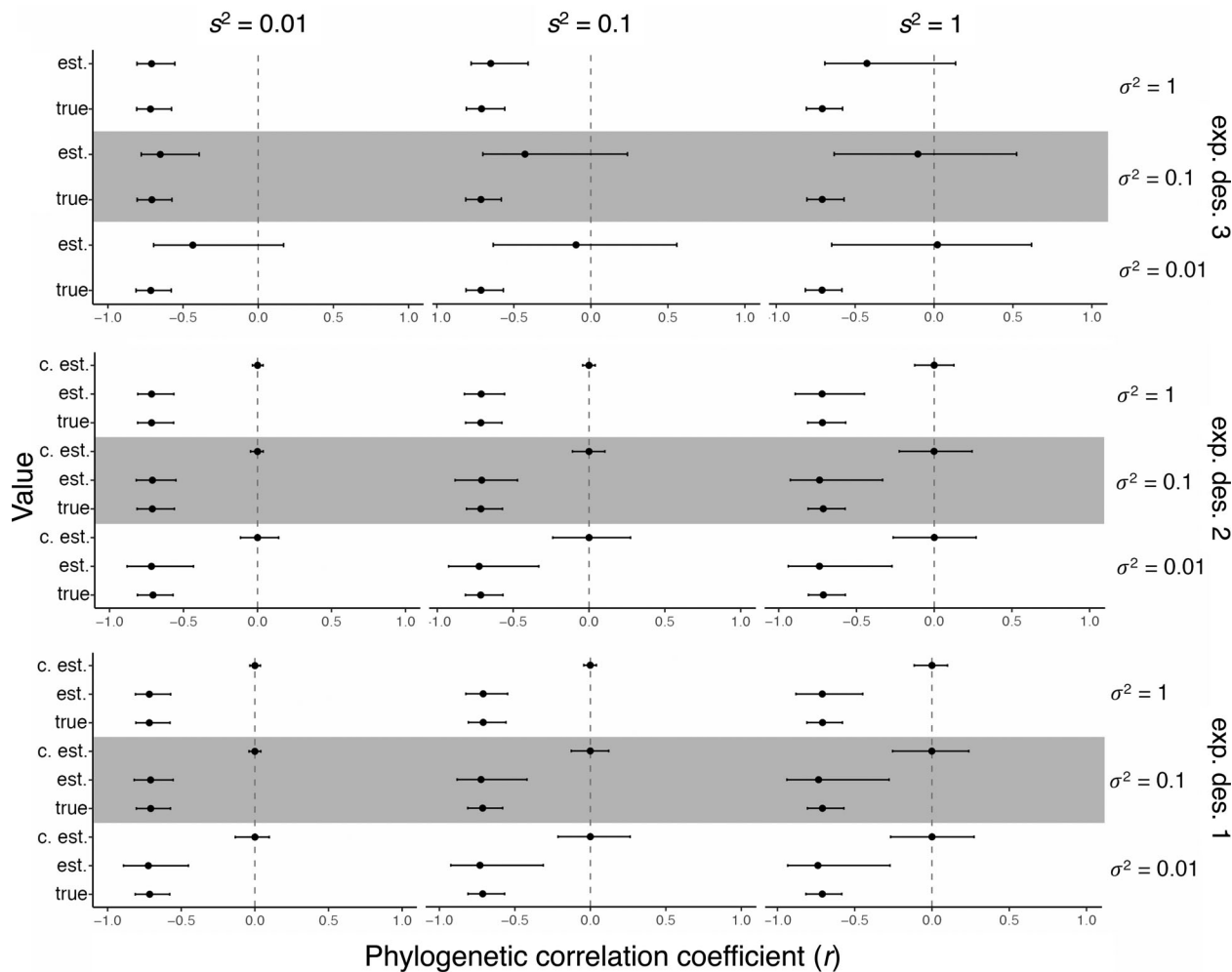
Figure 4. Summary of phylogenetic correlation coefficients between baseline phenotype and reaction norm slope from simulations in which plasticity evolved by Brownian motion via the Unlinked Phenotypes mechanism. Each panel shows the result of 1000 simulations. Each row shows the results of simulations with a different experimental design (Fig. 1): *Bottom*: Design 1. *Middle*: Design 2. *Top*: Design 3. Alternating shaded areas differentiate simulations with different Brownian motion rate parameters (σ^2). Each column contains simulations with a different error variance for phenotypic estimation (s^2). "true" denotes the true phylogenetic correlation values from the simulations. "est." denotes estimates of the true correlation coefficients given measurement error. "c. est" denotes estimated correlation coefficients after data adjustment to remove the effect of regression to the mean following our modified Kelly and Price (2005) correction. Corrected results are not given for Experimental Design 3 because the method cannot be applied to data collected with that design.

plasticity under any experimental design (Fig. 3, "true" values). The introduction of measurement error, however, can again lead to regression to the mean under Experimental Designs 1 and 2 (Fig. 3, "estimated" values). The degree to which regression to the mean is a problem depends on the magnitude of measurement error relative to the variance in true trait values among taxa as denoted by the Brownian motion rate parameter. If measurement error is large and/or phenotypic variation among taxa is low, regression to the mean consistently emerges with Experimental Designs 1 and 2. By contrast, if measurement error is small and

phenotypic variation among taxa is large, regression to the mean is much less of an issue (Fig. 3, "estimated" values).

Applying the Kelly and Price correction to data collected under Experimental Designs 1 and 2 can remove the effect of regression to the mean when it is present (Fig. 3, "corrected" values). Our results, however, also highlight the caveat that a Kelly and Price correction should only be applied if the initial analysis yields a significant negative relationship. When initial analyses yield no significant negative relationship, the Kelly and Price correction can itself create a spurious negative
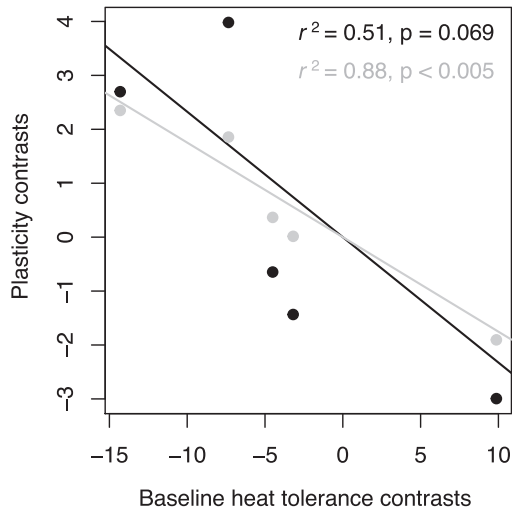
Figure 5. Nudibranch heat tolerance plasticity data from Armstrong et al. (2019). Phenotypes were measured using Experimental Design 2 (Fig. 1). Original data and analysis in gray. Reanalysis applying the modified Kelly and Price (2005) correction to remove regression to the mean effects in black.

relationship between baseline phenotype and plasticity. Experimental Design 3 again avoids the problem of regression to the mean due to the estimation of two baseline phenotypes (see above).

RESULTS UNDER NEUTRAL EVOLUTION OF PLASTICITY: UNLINKED PHENOTYPES MODEL

With the Unlinked Phenotypes model, the phenotypes in each environment evolve independently of one another, and the magnitude of plasticity emerges as a by-product. Under these conditions, a true negative association between baseline phenotype and plasticity consistently arises under Brownian motion evolution (Fig. 4, "true" values). In other words, there are conditions in which neutral evolution alone can produce a pattern consistent with genetic assimilation. This is in contrast to the most common explanations for genetic assimilation, which invoke some combination of selection and/or genetic constraint to explain why a reduction in plasticity evolves (West-Eberhard 2003; Pigliucci et al. 2006; Ehrenreich and Pfennig 2016).

The magnitude of experimental error has little effect on the correlation coefficients calculated using Experimental Designs 1 and 2 (Fig. 4, "estimate" values), contrary to the results with the Linked Phenotypes model. This is because phenotypic evolution via Brownian motion in the Unlinked Phenotypes model amounts to random independent increases or decreases in the phenotypes expressed in each environment, whereas experimental error leads to spurious random increases or decreases in the phenotypes measured in each environment. From a statistical standpoint, they are the same process.

When data from the Unlinked Phenotypes model are adjusted using the Kelly and Price correction, the negative relationship between baseline phenotype and plasticity disappears (Fig. 4, "corrected" values). In this case, however, the correction does not remove a spurious relationship due to regression to the mean. Instead, it removes a true relationship that resulted from a neutral evolutionary process. This occurs because of the conceptual foundation upon which the Kelly and Price correction is predicated: it assumes that a hallmark of regression to the mean is that phenotypic values between treatments (in this case, environments) within experimental units are uncorrelated with one another. This, of course, is the precise pattern that neutral evolution will produce if the phenotypes induced by plasticity are unlinked. This finding highlights that biological variation and error variation can be difficult to separate (Kelly and Price 2004; Hansen and Bartoszek 2012), and statistical correction for regression to the mean is not a panacea.

Correlation coefficients estimated using Experimental Design 3 can maintain the true negative relationship between baseline phenotype and plasticity, but only if the measurement error is not too large. As measurement error grows relative to the true phenotypic variance among taxa, estimated correlation coefficients move away from the true values and toward 0 (Fig. 4).

IMPLICATIONS

Our results make clear that regression to the mean is a serious statistical problem that can confound tests of genetic assimilation at macroevolutionary scales. Whether regression to the mean is an issue in any particular study will depend on the experimental design used. Unfortunately, the problematic experimental designs (our Designs 1 and 2) are those most often used to test for plasticity and genetic assimilation. This is certainly true within the field of thermal physiology, where tests of the Trade-Off Hypothesis, which states that greater thermal tolerance evolves via genetic assimilation (Heerwaarden and Kellerman 2020), are almost invariably based on data collected using Experimental Designs 1 and 2 (e.g., Armstrong et al. 2019). These studies typically find support for the Trade-off Hypothesis but do not account for regression to the mean.

Some evolutionary physiologists have used an alternative approach to test the Trade-off Hypothesis, in which thermal tolerance plasticity is correlated with habitat temperature (van Heerwaarden and Kellermann 2020). This approach avoids the regression to the mean problem, but it is important to recognize that modeling the association between plasticity and habitat tests a different hypothesis than modeling the association between plasticity and baseline phenotype. The latter directly pertains to the evolution of within-individual trait associations, trade-offs, and genetic assimilation. The former does not.

Fortunately, data collected using Experimental Designs 1 and 2 can be successfully adjusted to remove regression to the mean using the Kelly and Price method. For example, when plasticity is invariant among taxa, adjusting the data removes spurious relationships between baseline phenotype and plasticity due to measurement error (Fig. 3). The same holds under neutral evolution when there is linkage between the traits expressed in each environment (Fig. 3). The caveat is that adjusting data using the Kelly and Price method can incorrectly obfuscate a true negative relationship between baseline phenotype and plasticity if the traits expressed in each environment are unlinked (Fig. 4).

The need to adjust data can be avoided by using Experimental Design 3. This approach does not lead to spurious negative relationships between baseline phenotype and plasticity (Figs. 3, 4). It also recovers true negative relationships when they occur (Fig. 4). Unfortunately, this experimental design is significantly more labor intensive than Designs 1 and 2. It requires that we estimate phenotypes on more groups of individuals (3 vs. 2), and its reliability depends on the magnitude of measurement error in the estimation of phenotypic trait means for each environment, making large sample sizes crucial. There are issues with Experimental Design 3 as well. For example, when individuals have their phenotypes measured more than once, carryover effects of previous experience can influence results (O'Connor et al. 2014). Furthermore, for some taxa and traits, measuring the same individual twice will simply not be possible. One solution in such cases is to have three groups of individuals per taxa: two that have their phenotype measured under baseline conditions, and one that has phenotype measured under novel conditions. One baseline group is used to calculate plasticity, and the other is used as the baseline phenotype estimate. The key is that different baseline measurements are used for each.

Whether induced phenotypes are genetically linked influences null expectations for reaction norm evolution. Specifically, a macroevolutionary pattern consistent with genetic assimilation can arise under a neutral evolutionary process in the absence of selection and genetic linkage between phenotypes. The importance of genetic linkage for reaction norm evolution has long been appreciated. For example, the classic work of Via and Lande (1985) demonstrated that genetic covariance between induced phenotypes influences the trajectory of reaction norm evolution under selection. Unfortunately, genetic linkage between environmentally induced phenotypes is almost certainly trait, population, and environment dependent, and in many instances knowledge about linkage is completely lacking. Our results highlight the need for greater scrutiny of the genetic architecture of induced phenotypes.

Overall, we show that great care is necessary both in the design of macroevolutionary tests for genetic assimilation as well as in the interpretation of results. A lack of correlation between plasticity and baseline phenotype is not the correct null hypothesis in many cases due to the statistical phenomenon of regression to the mean. Once this is recognized, the potential solutions to the regression to the mean problem must be applied with care. Additionally, the expected null hypothesis under neutral evolution depends on the mechanism that underlies the expression of phenotypic plasticity. We hope that this analysis will be a helpful reference for others in the design and analysis of future studies of hypothesized genetic assimilation in phenotypically plastic traits.

REFERENCES
Armstrong, E.J., Tanner, R.L. & Stillman, J.H. (2019) High heat tolerance is negatively correlated with heat tolerance plasticity in nudibranch mollusks. *Physiological and Biochemical Zoology*, 92, 430–444.

Baker, J., Meade, A., Pagel, M. & Venditti, C. (2015) Adaptive evolution toward larger size in mammals. *Proc. Natl. Acad. Sci. USA*, 112, 5093–5098.

Barley, J.M., Cheng, B.S., Sasaki, M., Gignoux-Wolfsohn, S., Hays, C.G., Putnam, A.B., Sheth, S., Villeneuve, A.R. & Kelly, M. (2021) Limited plasticity in thermally tolerant ectotherm populations: evidence for a trade-off. *Proc. R. Soc. B Biol. Sci.*, 288, 20210765.

Barnett, A.G., Van Der Pols, J.C. & Dobson, A.J. (2005) Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*, 34, 215-220.

Chevin, L.-M., Lande, R. & Mace, G.M. (2010) Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS Biology*, 8, e1000357.

Chevin, L.-M., Gallet, R., Gomulkiewicz, R., Holt, R.D. & Fellous, S. (2013) Phenotypic plasticity in evolutionary rescue experiments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368, 20120089.

Comte, L. & Olden, J.D. (2017) Evolutionary and environmental determinants of freshwater fish thermal tolerance and plasticity. *Global Change Biology*, 23, 728-736.

Crispo, E. (2007) The Baldwin effect and genetic assimilation: revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity. *Evolution: International Journal of Organic Evolution*, 61, 2469-2479.

——— (2008) Modifying effects of phenotypic plasticity on interactions among natural selection, adaptation and gene flow. *Journal of Evolutionary Biology*, 21, 1460-1469.

De Jong, G. (1995) Phenotypic plasticity as a product of selection in a variable environment. *The American Naturalist*, 145, 493-512.

Deery, S.W., Rej, J.E., Haro, D. & Gunderson, A.R. (2021) Heat hardening in a pair of *Anolis* lizards: constraints, dynamics and ecological consequences. *Journal of Experimental Biology*, 224, jeb240994.

Ehrenreich, I.M. & Pfennig, D.W. (2016) Genetic assimilation: a review of its potential proximate causes and evolutionary consequences. *Annals of Botany*, 117, 769-779.

Esperk, T., Kjaersgaard, A., Walters, R.J., Berger, D. & Blanckenhorn, W.U. (2016) Plastic and evolutionary responses to heat stress in a temperate dung fly: negative correlation between basal and induced heat tolerance? *Journal of Evolutionary Biology*, 29, 900-915.

Ghalambor, C.K., McKay, J.K., Carroll, S.P. & Reznick, D.N. (2007) Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, 21, 394-407.

Ghalambor, C.K., Hoke, K.L., Ruell, E.W., Fischer, E.K., Reznick, D.N. & Hughes, K.A. (2015) Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature*, 525, 372-375.

Gunderson, A.R. & Stillman, J.H. (2015) Plasticity in thermal tolerance has limited potential to buffer ectotherms from global warming. *Proceedings of the Royal Society of London B: Biological Sciences*, 282, 20150401.

Gunderson, A.R., Dillon, M.E. & Stillman, J.H. (2017) Estimating the benefits of plasticity in ectotherm heat tolerance under natural thermal variability. *Functional Ecology*, 31, 1529-1539.

Gunter, H.M., Schneider, R.F., Karner, I., Sturmbauer, C. & Meyer, A. (2017) Molecular investigation of genetic assimilation during the rapid adaptive radiations of East African cichlid fishes. *Molecular Ecology*, 26, 6634-6653.

Hansen, T.F. & Bartoszek, K. (2012) Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. *Systematic Biology*, 61, 413-425.

Jackson, D.A. & Somers, K.M. (1991) The spectre of 'spurious' correlations. *Oecologia*, 86, 147-151.

Kelly, C. & Price, T.D. (2004) Comparative methods based on species mean values. *Mathematical Biosciences*, 187, 135-154.

——— (2005) Correcting for regression to the mean in behavior and ecology. *The American Naturalist*, 166, 700-707.

Kelly, M. (2019) Adaptation to climate change through genetic accommodation and assimilation of plastic phenotypes. *Philosophical Transactions of the Royal Society B*, 374, 20180176.

Kingsolver, J.G. & Buckley, L.B. (2017) Evolution of plasticity and adaptive responses to climate change along climate gradients. *Proceedings of the Royal Society B: Biological Sciences*, 284, 20170386.

Lande, R. (2009) Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of Evolutionary Biology*, 22, 1435-1446.

Martin, C.H., Crawford, J.E., Turner, B.J. & Simons, L.H. (2016) Diabolical survival in Death Valley: recent pupfish colonization, gene flow and genetic assimilation in the smallest species range on earth. *Proceedings of the Royal Society B: Biological Sciences*, 283, 20152334.

Morley, S., Peck, L., Sunday, J., Heiser, S. & Bates, A. (2019) Physiological acclimation and persistence of ectothermic species under extreme heat events. *Global Ecology and Biogeography*, 28, 1018-1037.

Murren, C.J., Auld, J.R., Callahan, H., Ghalambor, C.K., Handelsman, C.A., Heskel, M.A., Kingsolver, J. et al (2015) Constraints on the evolution

of phenotypic plasticity: limits and costs of phenotype and plasticity. *Heredity*, 115, 293-301.

O'Connor, C.M., Norris, D.R., Crossin, G.T. & Cooke, S.J. (2014) Biological carryover effects: linking common concepts and mechanisms in ecology and evolution. *Ecosphere*, 5, 1-11.

O'Neill, E., Davis, H.E. & MacMillan, H.A. (2021) A lack of repeatability creates the illusion of a trade-off between basal and plastic cold tolerance. *Proceedings of the Royal Society B*, 288, 20212121.

Pfennig, D.W., Wund, M.A., Snell-Rood, E.C., Cruickshank, T., Schlichting, C.D. & Moczek, A.P. (2010) Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution*, 25, 459-467.

Pigliucci, M., Murren, C.J. & Schlichting, C.D. (2006) Phenotypic plasticity and evolution by genetic assimilation. *Journal of Experimental Biology*, 209, 2362-2367.

Preston, J.C., Wooliver, R., Driscoll, H., Coughlin, A. & Sheth, S.N. (2022) Spatial variation in high temperature-regulated gene expression predicts evolution of plasticity with climate change in the scarlet monkeyflower. *Molecular Ecology*, 31, 1254-1268.

Price, T.D., Qvarnström, A. & Irwin, D.E. (2003) The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270, 1433-1440.

R Development Core Team (2021) R: a language and environment for statistical computing. The R Foundation for Statistical Computing, Vienna.

Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217-223.

Riddell, E.A., Odom, J.P., Damm, J.D. & Sears, M.W. (2018) Plasticity reveals hidden resistance to extinction under climate change in the global hotspot of salamander diversity. *Science Advances*, 4, eaar5471.

Rohr, J.R., Civitello, D.J., Cohen, J.M., Roznik, E.A., Sinervo, B. & Dell, A.I. (2018) The complex drivers of thermal acclimation and breadth in ectotherms. *Ecology Letters*, 21, 1425-1439.

Sasaki, M.C. & Dam, H.G. (2019) Integrating patterns of thermal tolerance and phenotypic plasticity with population genetics to improve understanding of vulnerability to warming in a widespread copepod. *Global Change Biology*, 25, 4147-4164.

——— (2021) Negative relationship between thermal tolerance and plasticity in tolerance emerges during experimental evolution in a widespread marine invertebrate. *Evolutionary Applications*, 14, 2114-2123.

Schlichting, C.D. & Pigliucci, M. (1998) Phenotypic evolution: a reaction norm perspective. Sinauer Associates, Sunderland, MA.

Schneider, R.F. & Meyer, A. (2017) How plasticity, genetic assimilation and cryptic genetic variation may contribute to adaptive radiations. *Molecular Ecology*, 26, 330-350.

Seebacher, F., White, C.R. & Franklin, C.E. (2015) Physiological plasticity increases resilience of ectothermic animals to climate change. *Nature Climate Change*, 5, 61-66.

Sikkink, K.L., Reynolds, R.M., Ituarte, C.M., Cresko, W.A. & Phillips, P.C. (2014) Rapid evolution of phenotypic plasticity and shifting thresholds of genetic assimilation in the nematode *Caenorhabditis remanei*. *G3: Genes, Genomes, Genetics*, 4, 1103-1112.

Somero, G. (2010) The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine 'winners' and 'losers'. *The Journal of Experimental Biology*, 213, 912-920.

Stillman, J.H. (2003) Acclimation capacity underlies susceptibility to climate change. *Science*, 301, 65-65.

van Heerwaarden, B. & Kellermann, V. (2020) Does plasticity trade off with basal heat tolerance? *Trends in Ecology & Evolution*, 35, 874-885.

Via, S. (1993) Adaptive phenotypic plasticity: target or by-product of selection in a variable environment? *The American Naturalist*, 142, 352-365.

Via, S. & Lande, R. (1985) Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution; Internation Journal of Organic Evolution*, 39, 505-522.

Waddington, C.H. (1953) Genetic assimilation of an acquired character. *Evolution; Internation Journal of Organic Evolution*, 7, 118-126.

West-Eberhard, M.J. (2003) Developmental plasticity and evolution. Oxford Univ. Press, Oxford, U.K.

Associate Editor: K. Montooth
Handling Editor: T. Chapman

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary Information